McKinsey&Company

MCKINSEY GLOBAL INSTITUTE

NOTES FROM THE AI FRONTIER

APPLYING AI FOR SOCIAL GOOD

DISCUSSION PAPER
DECEMBER 2018

Michael Chui | San Francisco
Martin Harryson | Silicon Valley
James Manyika | San Francisco
Roger Roberts | Silicon Valley
Rita Chung | Silicon Valley
Ashley van Heteren | Amsterdam
Pieter Nel | New York

MCKINSEY GLOBAL INSTITUTE

Since its founding in 1990, the McKinsey Global Institute (MGI) has sought to develop a deeper understanding of the evolving global economy. As the business and economics research arm of McKinsey & Company, MGI aims to provide leaders in the commercial, public, and social sectors with the facts and insights on which to base management and policy decisions.

MGI research combines the disciplines of economics and management, employing the analytical tools of economics with the insights of business leaders. Our "micro-to-macro" methodology examines microeconomic industry trends to better understand the broad macroeconomic forces affecting business strategy and public policy. MGI's in-depth reports have covered more than 20 countries and 30 industries. Current research focuses on six themes: productivity and growth, natural resources, labor markets, the evolution of global financial markets, the economic impact of technology and innovation, and urbanization. Recent reports have assessed the digital economy, the impact of AI and automation on employment, income inequality, the productivity puzzle, the economic benefits of tackling gender inequality, a new era of global competition, Chinese innovation, and digital and financial globalization.

MGI is led by three McKinsey & Company senior partners: Jacques Bughin, Jonathan Woetzel, and James Manyika, who also serves as the chairman of MGI. Michael Chui, Susan Lund, Anu Madgavkar, Jan Mischke, Sree Ramaswamy, and Jaana Remes are MGI partners, and Mekala Krishnan and Jeongmin Seong are MGI senior fellows.

Project teams are led by the MGI partners and a group of senior fellows, and include consultants from McKinsey offices around the world. These teams draw on McKinsey's global network of partners and industry and management experts. Advice and input to MGI research are provided by the MGI Council, members of which are also involved in MGI's research. MGI Council members are drawn from around the world and from various sectors and include Andrés Cadena, Sandrine Devillard, Tarek Elmasry, Katy George, Rajat Gupta, Eric Hazan, Acha Leke, Scott Nyquist, Gary Pinkus, Sven Smit, Oliver Tonby, and Eckart Windhagen. In addition, leading economists, including Nobel laureates, act as research advisers to MGI research.

The partners of McKinsey fund MGI's research; it is not commissioned by any business, government, or other institution. For further information about MGI and to download reports, please visit **www.mckinsey.com/mgi**.

Copyright © McKinsey & Company 2018

SUMMARY

NOTES FROM THE AI FRONTIER: APPLYING AI FOR SOCIAL GOOD

Artificial intelligence, while not a silver bullet, could contribute to the multi-pronged efforts to tackle some of the world's most challenging social problems. AI is already being leveraged in research to tackle societal "moon shot" challenges such as curing cancer and climate science. The focus of this paper is on other social benefit uses of AI that do not require scientific breakthroughs but that add to existing efforts to help individuals or groups in both advanced and developing economies who are experiencing challenges or crises and who often live beyond the reach of traditional or commercial solutions. We assess the AI capabilities that are currently most applicable for such challenges and identify domains where their deployment would be most powerful. We also identify limiting factors and risks to be addressed and mitigated if the social impact potential is to be realized.

WHAT'S INSIDE

1. Mapping AI use cases to domains of social good
Page 1

2. How AI capabilities can be used for societal benefit
Page 10

3. Six illustrative use cases
Page 18

4. Bottlenecks to overcome
Page 30

5. Risks to be managed
Page 35

6. Scaling up the use of AI for social good
Page 42

- Through an analysis of about 160 AI social impact use cases, we have identified and characterized ten domains where adding AI to the solution mix could have large-scale social impact. These range across all 17 of the United Nations Sustainable Development Goals and could potentially help hundreds of millions of people worldwide. Real-life examples show AI already being applied to some degree in about one-third of these use cases, ranging from helping blind people navigate their surroundings to aiding disaster relief efforts.
- Several AI capabilities, primarily in the categories of computer vision and natural language processing, are especially applicable to a wide range of societal challenges. As in the commercial sector, these capabilities are good at recognizing patterns from the types of data they use, particularly unstructured data rich in information, such as images, video, and text, and they are particularly effective at completing classification and prediction tasks. Structured deep learning, which applies deep learning techniques to traditional tabular data, is a third AI capability that has broad potential uses for social good. Deep learning applied to structured data can provide advantages over other analytical techniques because it can automate basic feature engineering and can be applied despite lower levels of domain expertise.
- These AI capabilities are especially pertinent in four large domains—health and hunger, education, security and justice, and equality and inclusion—where the potential usage frequency is high and where typically a large target population would be affected. In health, for example, AI-enabled wearable devices, which can already detect potential early signs of diabetes through heart rate sensor data with 85 percent accuracy, could potentially contribute to helping more than 400 million people afflicted by the disease worldwide if made sufficiently affordable. In education, more than 1.5 billion students could benefit from application of adaptive learning technology, which tailors content to students based on their abilities.

- Scaling up AI usage for social good will require overcoming some significant bottlenecks, especially around data accessibility and talent. In many cases, sensitive or monetizable data that could have societal applications are privately owned, or only available in commercial contexts where they have business value and must be purchased, and are not readily accessible to social or nongovernmental organizations. In other cases, bureaucratic inertia keeps data that could be used to enable solutions locked up, for example in government agencies. In most cases, the needed data have not been collected. Talent with high-level AI expertise able to improve upon AI capabilities and develop models is in short supply, at a time when competition for it from the for-profit sector is fierce. Deployment also often faces "last mile" implementation challenges even where data and technology maturity challenges are solved. While some of these challenges are nontechnical and common to most social good endeavors, others are tech-related: NGOs may lack the data scientists and translators needed to address the problem and interpret results and output from AI models accurately.
- Large-scale use of AI for social good entails risks that will need to be mitigated, and some tradeoffs to be made, to avoid hurting the very individuals the AI application was intended to help. AI's tools and techniques can be misused by authorities and others with access to them, and principles for their use will need to be established. Bias may be embedded in AI models or data sets that could amplify existing inequalities. Data privacy will need to be protected to prevent sensitive personal information from being made public and to comply with the law, and AI applications will need to be safe for human use. The continuing difficulty of making some AI-produced decisions transparent and explainable could also hamper its acceptance and use, especially for sensitive topics such as criminal justice. Solutions being developed to improve accuracy, including model validation techniques and "human in the loop" quality checks, could address some of these risks and concerns.
- Stakeholders from both the private and public sector have essential roles to play in ensuring that AI can achieve its potential for social good. Collectors and generators of data, whether governments or companies, could grant greater access to NGOs and others seeking to use the data for public service and could potentially be mandated to do so in certain cases. To resolve implementation issues will require many more data scientists or those with AI experience to help deploy AI solutions at scale. Capability building, including that funded through philanthropy, can help: talent shortages at this level can be overcome with a focus on accessible education opportunities such as online courses and freely available guides, as well as contributions of time by organizations such as technology companies that employ highly skilled AI talent. Indeed, finding solutions that apply AI to specific societal goals could be accelerated if technology players dedicated some of their resources and encouraged their AI experts to take on projects that benefit the common good.

The application of AI for societal benefit is an emerging topic and many research questions and issues remain unanswered. Our library of use cases is evolving and not comprehensive; while we expect to build on it, data about technological innovations and their potential applications are incomplete. Our hope is that this paper sparks further discussion about where AI capabilities can be applied for social good, and scaled up, so that their full societal potential can be realized.

1. MAPPING AI USE CASES TO DOMAINS OF SOCIAL GOOD

Artificial intelligence (AI), which for the purposes of this paper we use as shorthand to refer specifically to deep learning techniques, is increasingly moving out of the research lab and into the world of business.[1] Beyond its commercial uses, now increasingly widespread in mobile and other consumer applications, AI has noncommercial potential to do good. While AI is not a silver bullet or cure-all, the technology's powerful capabilities could be harnessed and added to the mix of approaches to address some of the biggest challenges of our age, from hunger and disease to climate change and disaster relief.[2]

Examples of where some of these capabilities are already being deployed illustrate how broad AI's impact could be. To cite just three: Planet Labs, an Earth-imaging Silicon Valley startup, partnered with Paul G. Allen Philanthropies and leading research scientists to create a global map of shallow-water coral reefs by applying object detection to satellite imagery in correlation with geospatial data. This map is used to monitor change over time and inform conservation interventions for the reef ecosystems that are under threat.[3] At Thorn, an international anti–human trafficking nonprofit organization, a combination of face detection and person identification, social network analysis, natural language processing, and analytics is being used to identify victims of sexual exploitation on the internet and dark web. Thorn works in collaboration with a group of technology companies, including Google, Microsoft, and Facebook, and has found a total of 5,791 child victims since 2016.[4] AI is also being used in the battle against cancer: researchers at the MIT Media Lab, for example, have applied reinforcement learning, a capability in which systems essentially learn by trial and error, in clinical trials with patients diagnosed with glioblastoma (the most aggressive form of brain cancer) to successfully reduce toxic chemotherapy and radiotherapy dosing. This example is particularly exciting as it shows capabilities still in development being applied to social good use cases; reducing chemotherapy doses helps improve quality of life of cancer patients and reduce the cost of their treatment. As further research continues to improve reinforcement learning, the practical applications of the solutions will extend beyond clinical trials to customization of patient treatment.[5]

In all, we have collected about 160 such social good use cases to date. They touch on some aspect of all 17 of the United Nations Sustainable Development Goals and potentially could help hundreds of millions of people worldwide. This use case library, which continues to grow and evolve, provides the basis for an in-depth examination of the domains where AI could be used and the applications that are likely to be the most impactful, as well as bottlenecks to impact and risks that will need to be addressed.

1 We recognize that the line of demarcation between artificial intelligence capabilities and other analytical capabilities is not universally shared, with different people holding different definitions, over time. In our use case library, we did estimate the potential for other analytical capabilities, including for the use of machine learning, as described in the Flint, Michigan, case in Chapter 3. Our use of "deep learning" refers to machine learning techniques on very large artificial (simulated) neural networks.

2 AI capabilities can be used for bad or malicious purposes as well as for social good. For a discussion of the ethics of AI, see Box 2, on page 36.

3 Andrew Zolli, *Planet, Paul G Allen Philanthropies, & leading scientists team up to map & monitor world's corals in unprecedented detail*, Planet, June 4, 2018, planet.com/pulse/planet-paul-g-allen-coral-map/.

4 Thorn's user surveys indicate that in the past two years, its "Spotlight" tool was used in 21,044 cases and identified 6,553 traffickers. wearethorn.org/spotlight/.

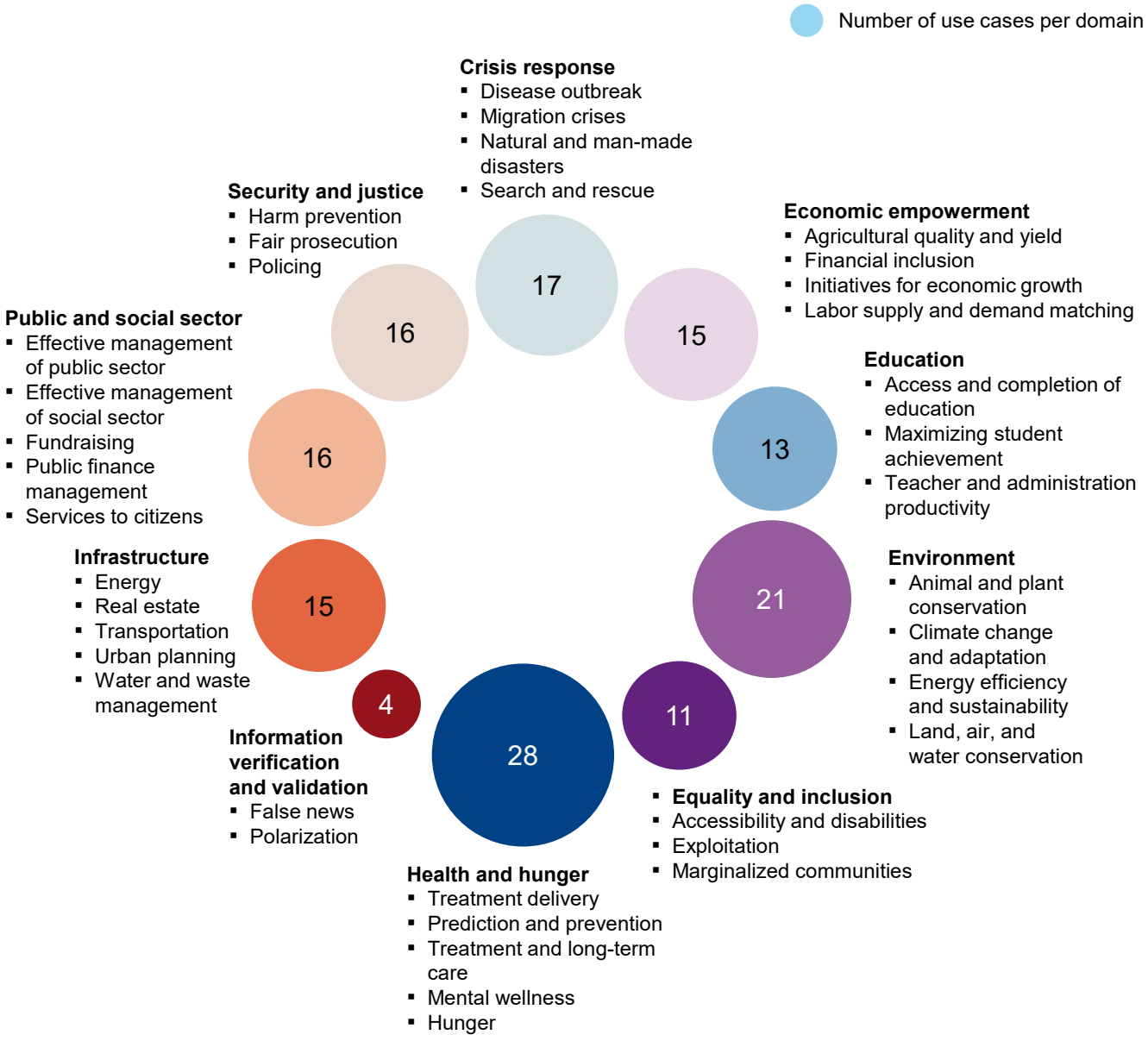
5 Rob Matheson, "Artificial intelligence model 'learns' from patient data to make cancer treatment less toxic," MIT News, August 9, 2018, news.mit.edu/2018/artificial-intelligence-model-learns-patient-data-cancer-treatment-less-toxic-0810.

AI'S POTENTIAL SOCIETAL IMPACT IS BROAD, BASED ON OUR MAPPING OF USE CASES TO DOMAINS

To build our library of use cases, which forms the basis of our analysis, we adopted a dual approach, both societal and technological (Exhibit 1). Each use case highlights a type of meaningful problem that can be solved by an AI capability or some combination of AI capabilities. To measure the relative potential of AI we used usage frequency as a proxy (see Box 1, "Building a library of AI use cases for social good to understand comparative relevance of AI across domains"). For about one-third of the use cases in our library to date, we identified an actual AI deployment in some form (Exhibit 2). Since many of these solutions are small test cases to determine feasibility, their functionality and scope of deployment often suggest that additional potential could be captured. For three-quarters of our use cases, we have seen deployment of solutions that employ some level of advanced analytics; most of these use cases, although not all, would further benefit from the use of AI applications.

Exhibit 1

Mapping domains to issue types and use cases in our library.

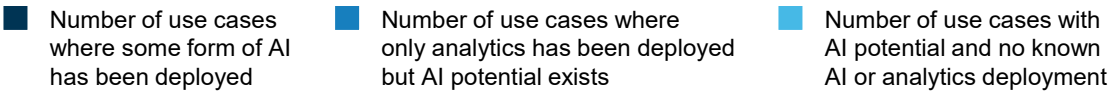


SOURCE: McKinsey Global Institute analysis

Exhibit 2

About one-third use cases in our library have been deployed in some form, leveraging AI capabilities.

USE CASE LIBRARY NOT EXHAUSTIVE


Number of use cases where some form of AI has been deployed
Number of use cases where only analytics has been deployed but AI potential exists
Number of use cases with AI potential and no known AI or analytics deployment

Social impact domain	Use case profile breakdown per domain	Remarks
Crisis response	7, 10, 17	This domain has high potential for AI use, although problems can be complex and developing the right AI solutions may take time.
Economic empowerment	5, 8, 2, 15	Existing use cases deploying AI typically related to agriculture; commercial market has supported AI solutions that could be adapted for societal good use cases. Many existing use cases that use analytics could benefit from structured deep learning, which is greatly underused today.
Education	5, 3, 5, 13	Most existing AI use cases employ natural language processing (NLP). For now, adaptive learning only leverages analytics.
Environment	12, 4, 5, 21	Research institutions and organizations working on AI use for social causes have supported AI deployment.
Equality and inclusion	8, 2, 1, 11	Many solutions in this domain rely on AI. Most use NLP and computer vision. Much room remains to raise the quality of solutions.
Health and hunger	10, 8, 10, 28	Existing cases that use AI are mainly focused on medical diagnoses, and deployment is not yet at scale.
Information verification and validation	3, 1, 4	High potential use of AI, but problem is complex and development of appropriate solutions may take time.
Infrastructure	1, 9, 5, 15	Most use cases in our library either do not have known case studies or use only analytics; type of problem to solve and data used mainly revolve around optimization using structured data.
Public and social sector	16	All use cases in this domain in our library have existing case studies and use only analytics, though NLP and structured deep learning would likely add value.
Security and justice	3, 10, 3, 16	Many potential AI solutions in this domain have not been implemented because of fear of negative repercussions. Existing use cases largely leverage analytics.

NOTE: Our library of about 160 use cases with societal impact is evolving and this chart should not be read as a comprehensive gauge of the potential application of AI or analytics capabilities.

SOURCE: McKinsey Global Institute analysis

Box 1. Building a library of use cases for social good to understand the comparative relevance of AI across domains

Our library of use cases, which forms the basis of our analysis, currently has about 160 use cases in ten social impact domains. To build the library, we adopted a two-pronged approach, both societal and technological (Exhibit 3).

Each use case highlights a type of meaningful problem that can be solved by an AI capability or some combination of AI capabilities. The problem to solve was given a broad enough definition so that similar solutions would be grouped together. Most domains have around 15 use cases, with two outlier domains of health and hunger (28 use cases) and information verification and validation (four use cases).

From a societal point of view, we sought to identify key problems that are known to the social-sector community and determine where AI could aid efforts to resolve them. From a technological point of view, we took a curated list of 18 AI capabilities and sought to identify which types of social problems they could best contribute to solving.

For each use case, we tried to identify at least one existing case study. Where none were identified, we worked iteratively with experts to identify gaps and added corresponding use cases to our library. To guide our thinking, we conducted interviews with some 80 AI practitioners, social entrepreneurs, domain experts, academics, and technology company executives.

The library is not comprehensive, but it nonetheless showcases a wide range of problems where AI can be applied for social good. As AI capabilities evolve and as technical and social impact practitioners continue to identify more ways in which these capabilities can be applied, we expect the library to grow.

Measuring usage frequency

To provide a rough (and admittedly imperfect) measure of the relative potential of AI, we employed usage frequency as a proxy for societal value. Unlike AI usage for commercial purposes, where the value is typically measured in dollars, social value is harder to measure across all domains and use cases using one metric. The cost of human suffering, whatever the cause, and the benefits of alleviating it, are impossible to precisely gauge and compare. Even comparisons using number of lives affected can quickly become meaningless both across domains and within them; for example, use cases that contribute to solving climate change issues can theoretically affect all seven billion people on the planet. However imperfect, the proxy of potential usage frequency of AI allows for comparisons between use cases individually or at an aggregate level across all domains, in terms of comparative magnitude of AI usefulness and impact.

To calculate AI usage frequency, we estimated the number of times that models trained using AI would be used in a year to predict an outcome. This quantitative approach provided a directional means to identify AI capabilities with higher potential to bring about social impact, and others where AI deployment would be useful but not as impactful. (A criterion for our use cases is that they reach a threshold of having "meaningful" societal value potential, as agreed by domain experts.)

AI usage frequency takes into account the number of individual cases for which a model would need to be run and how often the model would be run. This could be the number of lives affected in a use case and how often per year a model would be run on each individual.

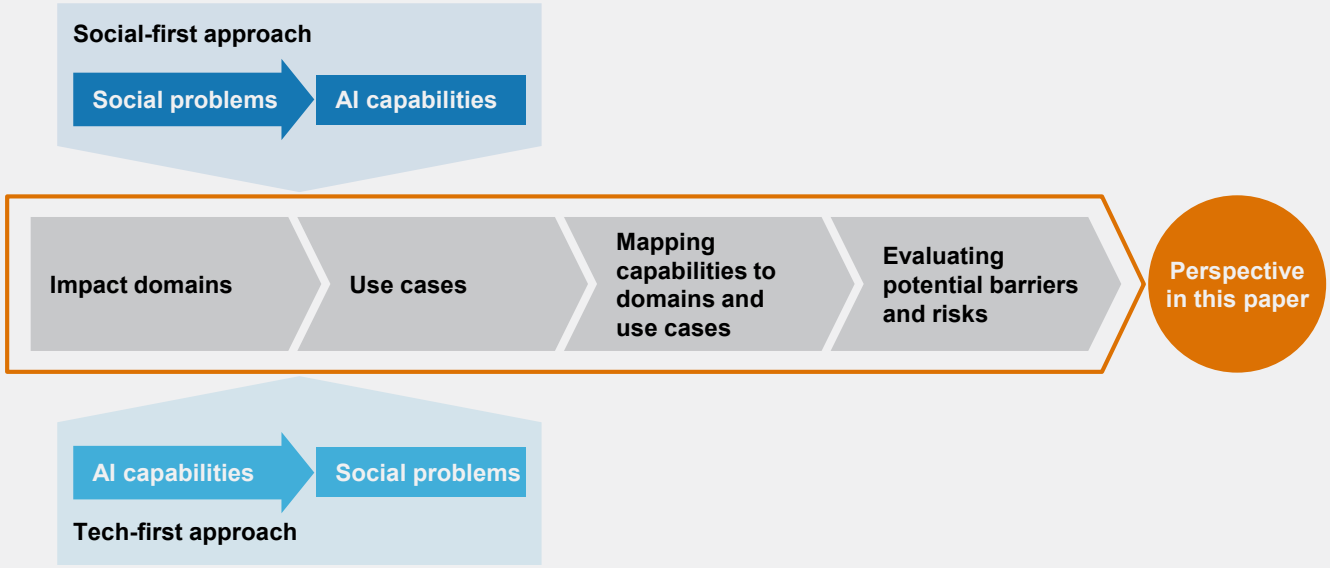
For example, in a use case on predicting students at risk of dropping out of school, the base is the number of K-12 students worldwide; the model in this case has to be run separately for each individual student approximately once per month to predict the likelihood that they will drop out. For an AI solution that uses a combination of capabilities including image classification, object detection, OCR, and emotion recognition to narrate the environment for the visually impaired, the base is the number of visually impaired people globally, and we estimate that it would be run nearly continuously, that is, once per minute for 16 active hours a day. The base number of individuals is not always the number of people. One example is from a use case where drones are used to detect poacher activity. Here, the metric we use is the 307 wildlife sanctuaries in the world, and we estimate that the system would be run once a minute for 12 hours a day (at night) when poachers are active.

Use cases can take in various data types and are often associated with more than one AI capability. We found that these vary significantly across social impact domains, based on our library. The heat map in Chapter 2 of this discussion paper focusing on AI usage frequency demonstrates this variety, highlighting the intersection of the domain or issue type and the specific AI capability.

Box 1. Building a library of use cases for social good to understand the comparative relevance of AI across domains (continued)

Exhibit 3

We built a library of use cases of AI for societal good using both social-first and tech-first approaches.



SOURCE: McKinsey Global Institute analysis

We grouped use cases into ten social impact domains based on examining and integrating taxonomies used by social-sector organizations, such as the AI for Good Foundation and the World Bank. Use cases within each domain are further grouped into two to five issue types.[6] The following is the list of the social impact domains we examined.

- **Crisis response.** Specific crisis-related challenges, such as responding to natural and man-made disasters in search and rescue missions and at times of disease outbreak. Examples of use cases with high potential usage frequency include using AI on satellite data to map and predict wildfire progression to optimize firefighter response. Drones with AI capabilities can also be used to find missing persons in wilderness areas.

- **Economic empowerment.** Opening access to economic resources and opportunities, including jobs, skills development, and market information, with an emphasis on currently vulnerable populations. For example, AI can be used for early detection of plant damage through low-altitude sensors, including smartphones and drones, to improve yield in small farms if farmers have access to technology; one project called FarmBeats is building edge-computing technology that could one day make data-driven farming accessible for even the poorest farmers.[7]

[6] AI for Good Foundation, ai4good.org/active-projects/. In July 2016, the World Bank introduced a new taxonomy of theme codes, projects.worldbank.org/theme.

[7] *GatesNotes*, "Can the Wi-Fi chip in your phone help feed the world?", blog entry by Bill Gates, October 9, 2018, gatesnotes.com/Development/FarmBeats.

- **Educational challenges.** These include maximizing student achievement and improving teacher productivity. For example, adaptive learning technology could be used to recommend content to students based on past success and engagement with the material. AI could also be used to detect student distress early, before a teacher has noticed.

- **Environmental challenges.** These include sustaining biodiversity and combating natural resource depletion, pollution, and climate change. For example, robots with AI capabilities can be used to sort recyclable material from waste. The Rainforest Connection, a Bay Area nonprofit, uses AI tools such as Google's TensorFlow in conservation efforts across the world. Its platform can detect illegal logging in vulnerable forest areas through analysis of audio sensor data.[8] Other applications include using satellite imagery to predict routes and behavior of illegal fishing vessels.[9]

- **Equality and inclusion.** Addressing equality, inclusion, and self-determination challenges, such as reducing or eliminating bias based on race, sexual orientation, religion, citizenship, and disabilities. One use case, based on work by Affectiva, which was spun out of the MIT Media Lab, and Autism Glass, a Stanford research project, involves use of AI to automate emotion recognition and provide social cues to help individuals along the autism spectrum interact in social environments.[10] Another example is the creation of an alternative identification verification system for individuals without traditional forms of ID, such as driver's licenses.

- **Health and hunger.** Addressing health and hunger challenges, including early-stage diagnosis and optimized food distribution. Researchers at the University of Heidelberg and Stanford University have created a disease detection AI system, using visual diagnosis of natural images such as images of skin lesions to determine if they are cancerous; the system outperformed professional dermatologists.[11] AI-enabled wearable devices, which can already detect potential early signs of diabetes through heart rate sensor data with 85 percent accuracy, could potentially help more than 400 million people worldwide afflicted by the disease if the devices could be made sufficiently affordable.[12] Other use cases include combining various types of alternative data sources such as geospatial data, social media data, telecommunications data, online search data, and vaccination data to help predict virus and disease transmission patterns, or using an AI solution to optimize food distribution networks in areas facing shortages and famine.

- **Information verification and validation.** The challenge of facilitating provision, validation, and recommendation of helpful, valuable, and reliable information to all. This domain differs from the others in that it focuses on filtering or counteracting content that could mislead and distort, including false and polarizing information disseminated through the relatively new channels of the internet and social media. Such content can have severely negative consequences, including the manipulation of election results and the mob killings in India and Mexico that have been triggered by false news

---

8 "What have we done so far?" Rainforest Connection, rfcx.org/home.

9 "Using satellite imagery to combat illegal fishing," *The Maritime Executive*, July 17, 2017.

10 affectiva.com and autismglass.stanford.edu. See also, David Talbot, "Digital summit: First emotion-reading apps for kids with autism," *MIT Technology Review*, June 9, 2014, https://www.technologyreview.com/s/528191/digital-summit-first-emotion-reading-apps-for-kids-with-autism/.

11 "Computer learns to detect skin cancer more accurately than doctors," *Guardian,* May 29, 2018.

12 Brandon Ballinger et al., *DeepHeart: Semi-supervised sequence learning for cardiovascular risk prediction*, 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA, February 2–7, 2018, aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16967/15916.

dissemination via messaging applications.[13] Use cases in this domain include actively presenting opposing views to ideologically isolated pockets in social media. Problems in this domain are both in the public interest and technically very complex. The main actors able to address them are most likely to be tech companies that own these platforms and technologies.

- **Infrastructure management.** Infrastructure challenges that could provide public good in the categories of energy, water and waste management, transportation, real estate, and urban planning. For example, traffic light networks can be optimized using real-time traffic camera data and Internet of Things sensors to maximize vehicle throughput. AI can also be used for predictive maintenance of public transportation systems such as trains and public infrastructure, including bridges, to identify potentially malfunctioning components.

- **Public and social sector management.** Initiatives that are related to the efficiency and effective management of public- and social-sector entities, including strong institutions, transparency, and financial management. For example, AI can be used to identify tax fraud using alternative data such as browsing data, retail data, and payments history. Another instance where AI can prove effective is in providing automated question answering via email to improve government interaction with citizens.

- **Security and justice.** Challenges in society that include harm prevention—both from crime and other physical dangers—as well as tracking criminals and mitigating bias of police forces. This domain focuses on security, policing, and criminal justice issues as a unique category adjacent to public-sector management. An example is using AI to create solutions that help firefighters determine safe paths through burning buildings using data from IoT devices.

THE SOCIAL DOMAINS COVERED BY OUR USE CASES RANGE ACROSS ALL 17 UNITED NATIONS SUSTAINABLE DEVELOPMENT GOALS

The United Nations Sustainable Development Goals (SDGs) are among the best-known and most frequently cited societal challenges, and our use cases map to all 17 of the goals, supporting some aspect of each one (Exhibit 4). The SDGs are contained in a 2030 Agenda for Sustainable Development adopted by all UN member states in 2015 which lays out strategies for ending poverty and other deprivations at the same time as improving health and education, reducing inequality, preserving the environment, and boosting economic growth, among other priorities.

Our use case library does not use the same taxonomy as the SDGs because their goals are not directly related to AI usage, unlike ours; about 20 of the cases in our library do not map to the SDGs at all.[14]

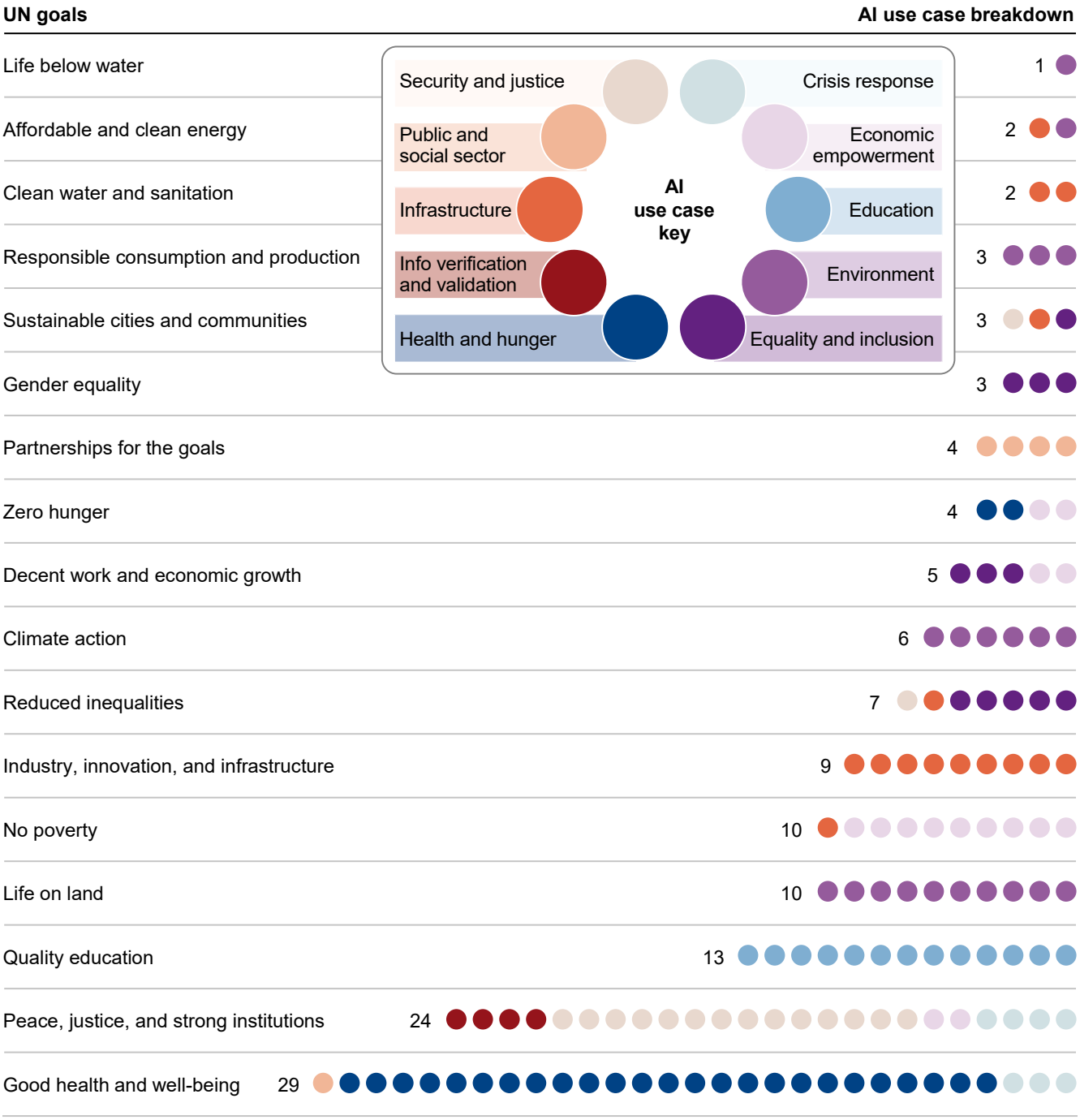
[13] Vindu Goel, Suhasini Raj, and Priyadarshini Ravichandran, "How WhatsApp leads mobs to murder in India," *New York Times*, July 18, 2018; Patrick J. McDonnell and Cecilia Sanchez, "When fake news kills: Lynchings in Mexico are linked to viral child-kidnap rumors," *Los Angeles Times*, September 21, 2018.

[14] For details of the 17 UN Sustainable Development Goals, see sustainabledevelopment.un.org/.

Exhibit 4

Use cases in our library support the UN Sustainable Development Goals.

UN Sustainable Development Goals[1]
Number of use cases



1 Chart is a partial list of use cases, as 21 of the 156 identified use cases do not target any of the UN's Sustainable Development Goals.
NOTE: This chart reflects the number and distribution of use cases and should not be read as a comprehensive evaluation of AI potential for each SDG; if an SDG has a low number of cases, that is a reflection of our library rather than of AI applicability to that SDG. The chart also does not reflect all use cases in the library, more than 20 of which do not map to any SDGs. These mainly focus on effective management in the public and social sectors, or belong to the issue types of disaster response and search and rescue in the crisis response domain.

SOURCE: United Nations; McKinsey Global Institute analysis

STAKEHOLDERS INCLUDING PUBLIC AND SOCIAL SECTOR ORGANIZATIONS ARE BUILDING THE MOMENTUM FOR SOCIAL BENEFIT APPLICATIONS

While AI solutions are not a first nor an obvious choice for many social-sector organizations, a variety of stakeholders in the social and public sectors, in universities, and in a number of commercial firms are starting to focus on the technologies in their work for social good. Two US-based NGOs already cited are examples: Thorn, co-founded by Hollywood stars Ashton Kutcher and Demi Moore, uses a range of computer vision and natural language processing (NLP) capabilities alongside analytics to identify and rescue victims of sexual trafficking, while Rainforest Connection uses sound detection from audio sensors to detect illegal logging in rainforests and then sends real-time alerts to partners on the ground. In Chapter 3, we describe six case studies in more detail. They show how a variety of organizations and academic institutions are taking the first steps to use AI solutions for social good. The organizations are not necessarily noncommercial; some are commercial companies either working on social good causes or using solutions that can have social value.

Beyond the examples we describe, our research found several organizations with AI expertise that are frequently cited in the sector for their work on applying AI to social good issues. They include AI4All, the AI for Good Foundation, DataKind, and Data Science for Social Good.[15] We expect that, with the right support from the broader AI for Social Good ecosystem, more organizations of all types will start to include AI as an option in their toolkit for solving social problems.

The geography of AI for social good is, for now, also quite narrow. While the use cases in our library provide coverage of AI being applied to social good around the world, most of the organizations behind these initiatives are based in the United States. This likely is a result of the concentration of high-level AI expertise, and it raises questions of how organizations with that expertise can better connect to communities around the world. As our library evolves, we anticipate seeing more use cases where not just deployment but also organization base and solution development location are in many different regions around the world.

[15] McKinsey & Company supports AI for social good causes through an initiative that sponsors projects and seeks to foster an ecosystem of joint efforts across NGOs, government organizations, technology partners, and others. Projects supported include ones to combat the spread of measles, help welfare case workers identify individuals who may need additional support in finding work, and help victims of human trafficking, including by training algorithms to search for signals in police and investigative databases.

2. HOW AI CAPABILITIES CAN BE USED FOR SOCIETAL BENEFIT

We identified 18 AI capabilities that could potentially be used for social benefit, of which 14 fall into three major categories: computer vision, natural language processing, and speech and audio processing. The remaining four, which we treated as stand-alone capabilities, include reinforcement learning, content generation, and structured deep learning. We also included a category for other analytics techniques (Exhibit 5). For the purposes of this paper, we use AI as shorthand specifically to refer to deep learning techniques that use artificial neural networks. While some machine learning techniques are sometimes referred to as AI, if they do not use deep learning we group them in this paper in the analytics category.

When we subsequently mapped these capabilities to domains (aggregating use cases) in a heat map, we found some interesting patterns (Exhibit 6). Making comparisons horizontally across each row highlights which capabilities could be particularly suitable for social good use cases. We see that a few are relevant across domains, a few are relevant to only specific domains, and some are barely used for now. We also made comparisons vertically to assess domains with regard to the potential use of AI. While many other insights can be drawn by mapping use cases, domains, and capabilities, we will describe some of our findings from this cross-domain view. The heat map is not a key from which to derive absolute impact, but a comparative index of potential and relevance of AI capabilities to certain domains and their use cases.

COMPUTER VISION, NATURAL LANGUAGE PROCESSING, AND STRUCTURED DEEP LEARNING HAVE BROAD POTENTIAL APPLICATION ACROSS DOMAINS

Based on AI usage frequency, several AI capabilities within the categories of computer vision and natural language processing and the capability of structured deep learning had the most widespread and greatest potential for social good application.

Image classification and object detection are powerful capabilities with multiple applications for social good

Within computer vision, the specific capabilities of image classification and object detection stand out for their potential applications for social good. Image classification is an AI capability that classifies an image or video clip into one of a set of predetermined classes; for example, image classification can tell you whether the image it reviews contains a cat or a dog. It can be used to solve problems such as mapping and predicting poverty levels of different neighborhoods based on nighttime luminosity and daytime imagery. Object detection is a capability that finds all instances of all trained classes of objects and reports their locations within an image; for example, object detection can find and highlight all locations where dogs appear in an image. This can be used to solve problems such as the detection of fires in satellite imagery.

These capabilities are often used in conjunction with one another. An example is when drones need computer vision capabilities to navigate a complex forest environment for search-and-rescue purposes. In this case, image classification may be used to distinguish normal ground cover from a footpath, guiding the drone's directional navigation, while object detection is used to circumvent obstacles such as trees. These capabilities are relevant in use cases where imagery is available and from which useful information can be extracted to solve a problem of social good.

Exhibit 5

Some AI capabilities could be used for societal benefit.

NOT EXHAUSTIVE

Capability category	Relative maturity	Capability	Example of problems the AI capability can solve
Computer vision	More developed	Person identification (image and video)	▪ Identifying a known missing child through publicly posted pictures and video (commonly referred to as facial recognition)
		Face detection (image and video)	▪ Detecting the presence of people in surveillance camera footage
		Image and video classification	▪ Identifying endangered animals in image and video for enhanced protection ▪ Detecting explicit content
		Near-duplicate or similar detection (images and video)	▪ Detecting hate-speech content for removal of image or video
		Object detection and localization (images and video)	▪ Detecting fires in satellite imagery
		Optical character and handwriting recognition (OCR, images)	▪ Digitizing hard-copy records for quicker patient health history search
		Tracking	▪ Tracking illegal fishing vessels via satellite imagery
	Developing	Emotion recognition (image and video)	▪ Measuring level of student engagement in classrooms
Speech and audio processing	More developed	Person identification (speech)	▪ Verifying individuals through mobile phone for inclusive banking access based on sound and pattern of voice
		Speech-to-text (audio)	▪ Real-time captioning for the deaf or people hard of hearing to facilitate live conversation
		Sound detection and recognition (audio)	▪ Identifying chain-saw sounds in rainforests for alerts on illegal logging activities
	Developing	Emotion recognition (speech)	▪ Assisting individuals on the autism spectrum who have difficulty in social interactions
Natural language processing	More developed	Person identification (text)	▪ Detecting a paper's author through handwriting analysis and identification of syntax patterns
		Language translation (text)	▪ Enabling larger distribution of online education services to underserved populations
		Other natural language processing (text)	▪ Identifying plagiarism in student assignments to enhance instructor productivity
	Developing	Sentiment analysis (text)	▪ Using automated review of public sentiment about specific topics to inform policy
		Language understanding	▪ Enabling chatbots that understand abstract concepts and ambiguous language, eg, ones that can do second-level, nuanced health screens
Content generation	Developing	Content generation	▪ Generating text and media (video, audio) content for educational purposes with quick production turnaround for wide distribution
Reinforcement learning	Developing	Reinforcement learning	▪ Large-scale and high-speed simulation modeling, for example in drug trials, doing millions of simulations to determine best treatment for breast cancer in population with a specific genetic makeup
Deep learning on structured data	More developed	Structured deep learning	▪ Identifying tax fraud and underreporting of income based on tax return data
Analytics	More developed	Analytics	▪ Any analytics technique not involving deep learning, eg, for optimization, journey mapping, network analysis

SOURCE: McKinsey Global Institute analysis

Exhibit 6

Mapping usage frequency of AI capabilities to ten social impact domains identifies patterns of the relevance and applicability of AI for social good.

Usage frequency of AI capability for each domain

Capability ▼ / Domain ▶	Equality and inclusion	Education	Health and hunger	Security and justice	Info verification and validation	Crisis response	Economic empowerment	Public and social sector	Environment	Infrastructure
Deep learning on structured data	●	●	●	●	●	●	●	●	●	●
Natural language processing[2]	●	●	●	●	●	●	●	●	●	
Image and video classification	●	●	●	●	●	●	●		●	●
Object detection and localization	●		●	●	●	●	●		●	●
Language understanding	●	●	●	●	●	●	●	●		
Sound detection and recognition	●	●	●	●		●	●		●	●
Sentiment analysis	●	●	●	●	●	●		●		
Language translation	●	●	●	●		●		●		
Face detection	●	●		●	●	●				
Tracking		●	●	●		●			●	
Emotion recognition	●	●	●		●					
Person identification	●	●		●		●				
Optical character and handwriting recognition	●	●		●						●
Speech-to-text	●	●	●							
Content generation	●		●					●		
Reinforcement learning			●							
Near-duplicate or similar detection				●						

Usage frequency of analytics capability for each domain

Analytics, eg, optimization, network analysis	●	●	●	●	●	●	●	●	●	●

1 Log base 10 scale. Deployment frequency capped at once per hour per year to prevent skewing; capping affected only a small number of use cases.
2 Excluding sentiment analysis, speech-to-text, language understanding, and translation.
NOTE: Our library of about 160 use cases with societal impact is evolving and this heatmap should not be read as a comprehensive gauge of the potential application of AI or analytics capabilities. Usage frequency estimates the number of times that models trained using AI would be used in a year to predict an outcome.

SOURCE: McKinsey Global Institute analysis

Some of these use cases consist of tasks individual humans could potentially accomplish, but where the required number of instances is so large that it exceeds human capacity, such as finding flooded and unusable roads across a large area after a hurricane. In other cases, an AI system can perform with greater accuracy than a human (often by processing more information), for example early identification of plant disease to prevent infection of the entire crop, which can be devastating to small-scale farmers. In some cases, a machine may be able to pick up key features of an image where a human would not or could not. One example is the use of person identification to identify victims of exploitation.

Computer vision capabilities such as person identification, face detection, and emotion recognition are relevant only in select domains and use cases, including for crisis response, security, equality, and education—but where they are relevant, their impact is great. In these use cases, the common theme is the need to identify individuals, which is most easily done through analysis of images. An example is the use of face detection on surveillance footage to detect the presence of escaped criminals in a specific area.[16] In education, emotion recognition on video or image data can be helpful in determining which specific students need extra attention and help. As we discuss elsewhere, these capabilities also introduce significant risks, for example around privacy, that must be understood and managed.

Natural language processing can be applied for societal impact where language and communication barriers are a roadblock

Some aspects of natural language processing, including sentiment analysis, language translation, and language understanding, also stand out as applicable to a wide range of domains and use cases. Natural language processing is most useful in domains where information is commonly stored in unstructured textual form, such as incident reports, health records, newspaper articles, and social media posts such as tweets.

As with computer vision–based methods, in some cases a human may be able to perform the task with greater accuracy than a trained machine learning model. However, the speed of "good enough" automated systems can enable meaningful scale efficiencies—for example, providing automated answers to questions asked by citizens through email. There are also cases where AI models could outperform humans in effectiveness, especially in situations that require processing and analyzing vast amounts of information quickly. Examples include monitoring disease outbreaks by analyzing tweets sent in multiple local languages.

As with computer vision capabilities, some language-related capabilities are not used in many domains, but they add substantial incremental value where they are used. For instance, language translation provides value in domains where language and communication barriers are a major roadblock, for example when migrant populations have difficulty communicating in the local language. In health, we often see challenges in treatment due to language barriers between doctors and patients. Language translation can also be used in education, to help students learn new languages.

Some capabilities, or combinations of capabilities, can open possibilities for the target population that would not otherwise exist—especially in use cases that involve understanding the natural environment through interpretation of vision, sound, and speech. For example, real-time description of one's environment could be game-changing in helping blind people navigate their surroundings. Another example is the use of AI to help educate children who are on the autism spectrum. Although professional therapists have proven effective in creating behavioral learning plans for children with autism spectrum disorder,

[16] Face detection applications detect the presence of people in an image or video frame. This is not the same as facial recognition, which is used to identify individuals by their features.

waiting lists for therapy can be long.[17] AI tools, primarily using emotion recognition and face detection capabilities, can increase access to this education by providing cues to help children identify and ultimately learn facial expressions among their family members and friends.

Structured deep learning analyzes traditional tabular data sets, which are often accessible for societal impact uses

The potential of structured deep learning across domains that we find in our use case library is perhaps more surprising because it is not commonly used today. This capability involves analysis of traditional tabular data sets using deep learning. It can contribute to solving problems ranging from identifying fraud based on tax return data to finding patterns of insights in electronic health records that would be very hard for humans to discover.

Structured deep learning (SDL) has been gaining momentum in the commercial sector, and we expect to see that trend spill over into solutions built for social good use cases, particularly given the abundance of tabular data in the public and social sectors. The advantage of SDL solutions is that they reduce the need for domain expertise or for an innate understanding of the data by automating aspects of basic feature engineering, the process of selecting and encoding the features in the data that will be most relevant. This could make SDL easier to use compared with other analytics models, especially for groups like NGOs that have limited resources and talent.

For now, the value of using structured deep learning is only starting to emerge. For example, in commercial applications, Instacart has successfully used SDL to enable its professional shoppers to efficiently navigate stores, reducing shopping time by minutes per delivery—an important efficiency gain.[18] In another example, Pinterest developed an SDL system to surface relevant recommendations that led to a 5 percent increase in its "related pin" engagement, a funnel to money-making features such as direct advertising and monetizable pins.[19] A third example is the use of trip data to train structured deep learning systems that predict taxi trajectories. In a Kaggle competition, this solution outperformed all non–deep learning solutions to win first place, increasing the efficiency of electronic taxi dispatching systems and helping optimize public transport.[20]

While our research has so far not identified actual deployment of SDL for social good in the world today, our use case library suggests that it has considerable potential to be used for noncommercial purposes across a range of domains because tabular data exists in relative abundance in almost every domain. It is by far the most common data type in the public- and social-sector management and infrastructure domains. For example, it is easy to see potential for public-sector agencies to deploy types of SDL architecture similar to Instacart's shopper navigation solution to optimize deployment of emergency vehicles, police, and other agents for greater workforce efficiency than can be gained through optimization methods using traditional analytics. Other potential examples include using structured deep learning in education to help recommend educational and vocational opportunities using academic history and financial information. In the equality and inclusion domain, it could be used to automatically enroll individuals on welfare in programs of which they were not aware, but for which they were qualified. In the health and hunger domain, too, broad applications, such as increasing patient support through identifying drivers of hospital performance, have potential.

---

17 "Autism therapy wait list changes 'a difficult process,' minister acknowledges," *CBC News*, April 2, 2016, cbc.ca/news/canada/ottawa/autism-wait-list-ontario-minister-1.3517508.

18 Jeremy Stanley, "Deep learning with emojis (not math)," *Medium*, March 29, 2017, tech.instacart.com/deep-learning-with-emojis-not-math-660ba1ad6cdc.

19 Kevin Ma, "Applying deep learning to Related Pins," *Medium*, January 12, 2017, medium.com/the-graph/applying-deep-learning-to-related-pins-a6fee3c92f5e.

20 *No free hunch*, "Taxi trajectory winners' interview: 1st place, team?," blog entry by Kaggle Team, July 27, 2015, blog.kaggle.com/2015/07/27/taxi-trajectory-winners-interview-1st-place-team-%F0%9F%9A%95/.

OTHER CAPABILITIES HAVE SOCIAL POTENTIAL, INCLUDING SOUND RECOGNITION, REINFORCEMENT LEARNING, AND ADVANCED ANALYTICS

Beyond these three capabilities, our use case library suggests that other capabilities, including both developing AI ones and more established advanced analytics, have potential applications for social benefit.

Sound detection and recognition could become increasingly relevant as auditory sensors are deployed for use cases across domains

Sound detection and recognition is also a fairly mature capability that has widespread reach given the prevalence of audio in the natural world. It is a common AI capability that can be used across many domains. As more auditory sensors are deployed, the scope of this capability will increase correspondingly. Today, sound detection is an important capability for use cases such as the exposure of illegal logging, diagnosis of medical or neurological conditions such as Parkinson's disease, predictive maintenance of public transportation systems, and providing adaptive learning to students.

Reinforcement learning and content generation, while nascent capabilities, have potential for social use

Reinforcement learning and content generation are relatively young capabilities and are not particularly prevalent in our use case library, but they do feature in a few cases. Reinforcement learning is a capability in which systems are trained by receiving virtual "rewards" or "punishments," essentially learning by trial and error. It was the primary technique used in the breakthrough AlphaGo program, which became the first program to beat a human professional Go player. It can be used, for example, to train models that recommend precision medicine–based treatments for individual patients. Researchers from the MIT Media Lab created a model that explores previous oncology drug regimens and at each monthly checkpoint determines updates that need to be made to patient doses, with the end goal of shrinking tumor size.[21]

Recent progress in visual content generation has been made using generative adversarial networks (GANs), a neural network architecture. Given their ability to mimic data sets, GANs are particularly effective at augmenting existing data sets.[22] GANs can also carry out a number of image-related tasks, including transforming visuals into a different style, improving resolution of low-resolution images, and repairing holes in images. The ability to generate language content is still developing and will need more improvement before the potential to generate meaningful text, such as in chatbot responses for mental wellness and medical consultations, can be fully captured.

Advanced analytics can be a more time- and cost-effective solution than AI for some use cases

Some use cases in our library are better suited to analytics techniques other than those that involve deep learning, since analytics are often less complex to deploy than those involving AI techniques. Moreover, for certain tasks, other analytical techniques can be better suited than deep learning. For example, in cases where there is a premium on explainability, decision tree–based models can often be more easily understood by humans. A financial institution might deploy into production tree-based models for underwriting, despite having parameters that were set based on experiments with deep learning.

---

[21] Rob Matheson, "Artificial intelligence model 'learns' from patient data to make cancer treatment less toxic," MIT News, August 9, 2018, news.mit.edu/2018/artificial-intelligence-model-learns-patient-data-cancer-treatment-less-toxic-0810.

[22] GANs operate by use of two neural networks: the "generator" creates new instances of data, while the "discriminator" evaluates their authenticity.

The full range of analytics techniques is likely to continue playing a key role because structured deep learning is only just gaining momentum and has not shown marked improvement over traditional methods for optimization problems, except in select commercial use cases. One example is the use case of optimizing public-sector agent deployment (for example, firefighters and EMTs). While SDL may surpass the use of analytics for this use case in the next few years, it is important to note that analytics techniques have a lower barrier to deployment and may thus be a preferable solution for an organization, given the available resources and technical capabilities.

THE VERTICAL VIEW OF DOMAIN COMPARISONS BRINGS OUT NUANCES IN THE PREVALENCE OF DIFFERENT DATA MODALITIES

In our use, modality refers to the type of data an AI system uses—for example, data from light sensors (that is, image and video) or from audio sensors. A multimodal AI solution is one that takes in different types of data inputs. Adaptive learning, assisting those with disabilities, assisting firefighters in navigating paths in a fire, and using images, audio, and even geospatial data to identify extremist online content for removal are all examples of multimodal applications for social good. Modality is a potentially important issue, especially for social good, because multimodal AI solutions are highly complex and may take more time and more resources to build (but might also have higher performance) than unimodal solutions.

Across domains, a subset of use cases (only 28 percent of the total) require multimodality to unlock maximum societal value. Three-quarters of them are concentrated in the same five domains where many AI capabilities are applicable: crisis response, education, equality and inclusion, health and hunger, and security and justice (Exhibit 7). Overall, among these five domains, only 38 percent of use cases are multimodal; the remaining 62 percent either use only a single source of unstructured data and perhaps structured data or use only structured data.

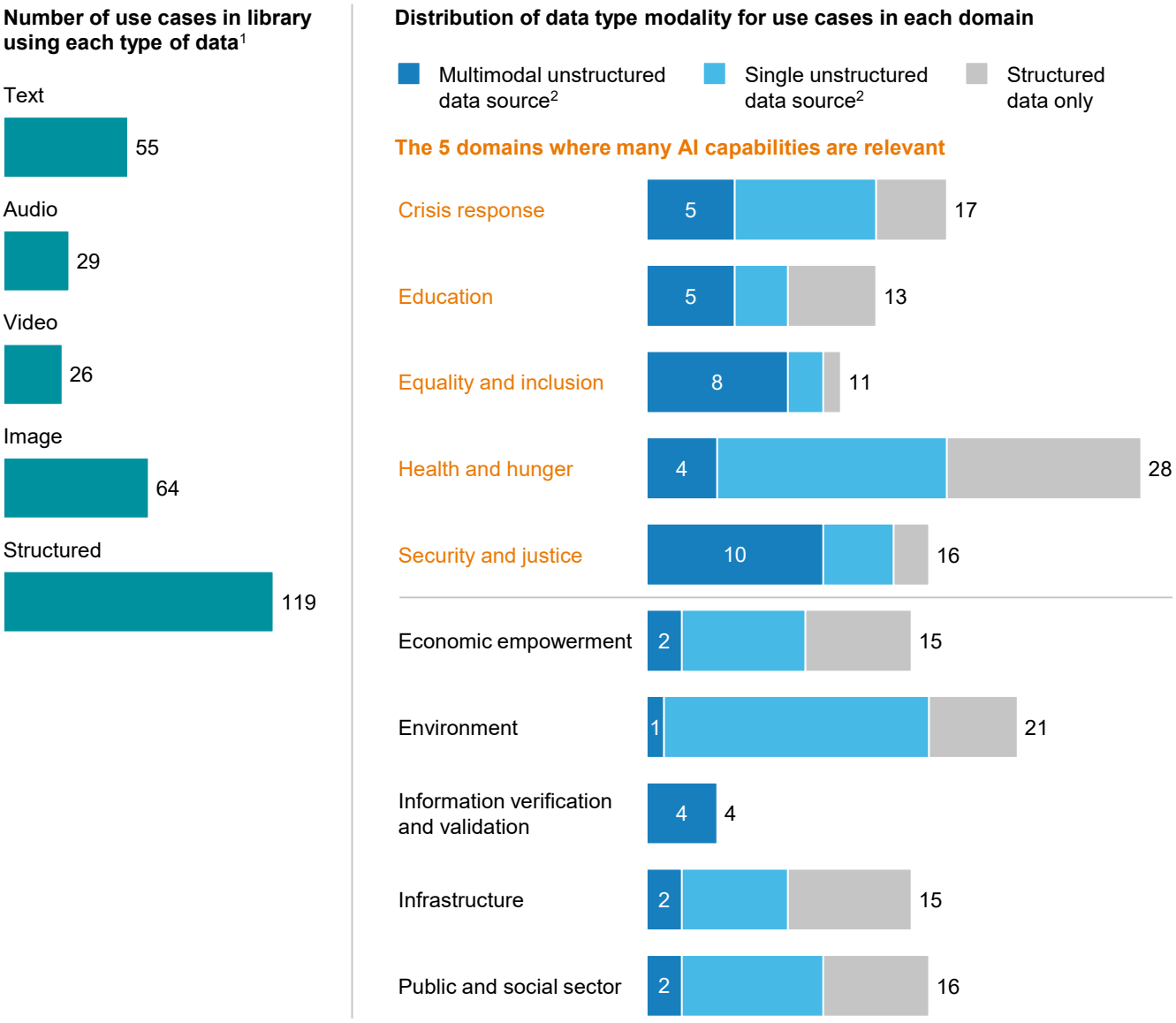
These domains split into two groups. For crisis response, education, and health and hunger, we find many AI capabilities that can be deployed because their use cases require a wide variety of types of data input, and most use only one type of unstructured data. This results in many AI capabilities being applicable across these domains.

For the equality and inclusion and the security and justice domains, however, we find a high percentage of multimodal use cases (73 percent and 63 percent, respectively), due to an emphasis on natural environment and human behavior understanding in these domains. Use cases can include, for example, narrating the environment for the blind and decreasing search time for criminal suspects using audio and video surveillance footage, which typically require multimodality. Such solutions will be more complex to build, and the impact in these two domains may thus not be realized as quickly as in other domains.

Exhibit 7

More than 70 percent of use cases require only one modality. Five domains for which many AI capabilities are relevant have the biggest share of multimodal use cases.

Modality is the way something is experienced, eg, seen (image and video), heard (audio)



1 Aggregate sum of use cases in this chart exceeds 156 because use cases can employ more than one type of data.
2 The category "Single unstructured data source" represents use cases that employ only one type of unstructured data (eg, text, image, video, or audio) but may or may not also make use of structured data; for "Multimodal unstructured data source," two or more types of unstructured data are required by the use cases, and structured data may be used as well.
NOTE: Our library of about 160 use cases with societal impact is evolving and this chart should not be read as a comprehensive gauge of the potential application of AI or analytics capabilities.

SOURCE: McKinsey Global Institute analysis

3. SIX ILLUSTRATIVE USE CASES

AI will be deployed in different ways depending on the domain, capability, barriers, and risk profiles of specific use cases. To illustrate the breadth of areas where AI could be applied for social good, this section goes into more specific detail about six cases. While these are just a specific set of examples, they highlight the range of cases, the different capabilities that could be used, and the potential impact. Several of the cases use capabilities in the computer vision category, including helping the visually impaired globally (number 1), tracking wildlife poachers in South Africa (number 4), and responding to crises such as Hurricane Harvey in Houston (number 6). The fifth case, focused on the prediction of houses in Flint, Michigan, that may have lead pipes, uses machine learning techniques. While these are often considered to be AI, they are classified under the umbrella of analytics for this paper, since we use a narrower definition of AI as deep learning. This use case example underscores how AI is one solution among others; machine learning and analytics solutions are also applicable and sometimes better suited for some social problems than deep learning.

1. Using computer vision to help the visually impaired navigate their environment

An estimated 250 million people worldwide have moderate to severe visual impairment, and about 90 percent of them live in developing nations (Exhibit 8). Not being able to see is a major impediment to work; about 70 percent of working age adults with visual impairment in the United States are unemployed. The World Health Organization estimates that the global economic impact of unaccommodated blindness and low vision worldwide exceeds $42 billion.[23]

Better medical care, nutrition, and other measures can help prevent blindness or ease the suffering of those afflicted by it. Alongside these solutions, AI has the potential to make life easier for the visually impaired by providing navigational assistance via smartphone. AI's computer vision capabilities can identify objects and read text, converting handwriting or printed text to digital text that can then be spoken aloud. Existing mobile applications include Microsoft's Seeing AI, which is accessible to users in 70 countries around the world and free of charge. Reviews of the application indicate that it has very large daily applicability.[24] Similar solutions include the OrCam MyEye camera, which is mounted on standard glasses and converts what is seen into spoken output. The device is easily portable and operates without the need for a smartphone but carries a $3,000 price tag.

While advances in computer vision have improved scene description technology, this remains a developing capability. As it improves, it will provide visually impaired people with a richer and deeper understanding of their environment, including recognizing other people and describing colors, for example.

Limited access to technology in emerging economies is a potential barrier that will need to be overcome

One of the most significant barriers to AI in this area is "last mile" implementation in the form of access to technology. Most visually impaired people live in emerging economies where internet access and bandwidth may be lacking, limiting the ability of the user to send a picture to the cloud for processing. Many people do not own the smartphone needed for existing applications; smartphone penetration globally is below 40 percent.[25] Building partnerships with NGOs and governments to provide basic technology access to individuals in poor communities could help address this limitation.

Talent is also a potential bottleneck. Developing AI applications to tailor them more closely to the needs of the visually impaired in various regions, including with more detailed description of the surrounding environment, may require the efforts of high-level AI researchers over a period of years.

[23] The World Health Organization and more than 20 international NGOs in 1999 launched a global initiative, Vision 2020, that aims to eliminate avoidable blindness by the year 2020. The initiative provides guidance and technical and resource support to countries that have formally adopted its agenda. See iapb.org/vision-2020/.

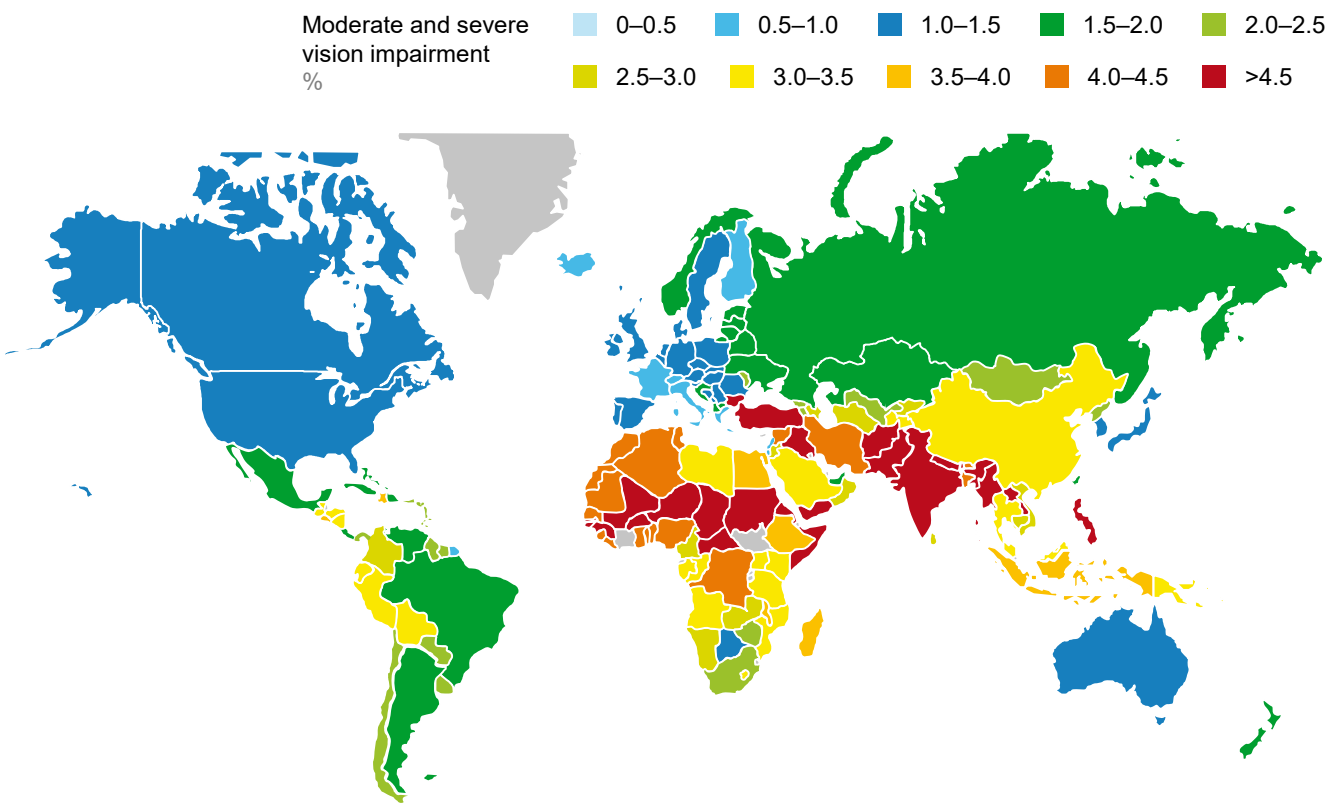
[24] See for example, Steven Kelley, "Seeing AI: Artificial Intelligence for blind and visually impaired use," VisionAware, American Foundation for the Blind, visionaware.org/info/everyday-living/helpful-products/using-apps/seeing-ai-app/1235.

[25] *The mobile economy 2018*, GSMA, 2018, gsma.com/mobileeconomy/wp-content/uploads/2018/05/The-Mobile-Economy-2018.pdf.

Exhibit 8

Most visually impaired people in the world live in emerging economies.

2015



SOURCE: The Vision Loss Expert Group; McKinsey Global Institute analysis

Among the risks, data privacy looms large. Users of such a mobile application would capture images of the environment that often include other individuals or private data, for example pictures of credit cards, and may also rely on training data that could include private information. Since these images could be automatically uploaded to the cloud, application developers will need to take steps toward protecting data privacy. Anonymized storage of data provided through the app may ensure security and privacy in the event of a data breach, although there are some doubts in the industry about the effectiveness of anonymization. Another strategy is to rely more heavily on edge computing and process the data on the device where it is being generated (the user's smartphone) instead of in a centralized data-processing facility (the cloud). This minimizes the chances that a central data storage breach will affect a user whose information is mainly stored in a distributed fashion.

2. Using computer vision to diagnose skin cancer from mobile phone photos

One in three cancer diagnoses is for skin cancer. When detected at an early stage, skin cancer survival rates can be as high as 97 percent, but they drop as low as 14 percent with late-stage detection.[26] Detection today is largely done by dermatologists looking at moles with the naked eye or a dermoscope. Residents of rural communities without dermatologists in the surrounding area are consequently at particular risk of late detection.

Some experiments suggest that AI can diagnose skin cancer with greater accuracy than human dermatologists; in a pilot, AI classification of skin lesions as melanomas versus benign beat classification by 58 dermatologists. AI diagnosed cancerous moles with 95 percent accuracy, while the dermatologists' accuracy was 86 percent.[27] This

[26] Taylor Kubota, "Deep learning algorithm does as well as dermatologists in identifying skin cancer, *Stanford News*, January 25, 2017, https://news.stanford.edu/2017/01/25/artificial-intelligence-used-identify-skin-cancer/.

[27] Thirty of the dermatologists were deemed experts, defined as having had more than five years of experience; 11 were deemed "skilled," with between two and five years' experience; and 17 were "beginners," with less than two years' experience. H. A. Haenssle et al., "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists, *Annals of Oncology*, August 2018, Volume 29, Issue 8, pp. 1836–1842, https://www.ncbi.nlm.nih.gov/pubmed/29846502.

creates the potential for a mobile application using image recognition to make testing accessible to all, including rural communities around the world that do not have easy access to dermatologists.[28]

This AI solution would use object detection and image classification of skin cancer images. A convolutional neural network system could be leveraged to build detection of cancerous versus regular skin lesions. The model would be trained using image data at a raw pixel level with associated cancer versus non-cancer tags. Training would require the aggregation of data sets with thousands of such images and diagnostic information. During deployment, images could potentially be submitted through an internet-based mobile app or through MMS to maximize the user base.

Further testing and development will be needed to ensure that the benefits are realized on a broad scale

While the technology for deployment already exists and has been piloted, a solution for mobile usage would need to be packaged and further tested before it could be used and scaled. This can be accomplished fairly easily by high-level engineering talent. However, to improve diagnosis of additional skin cancers not included in the initial model and continue to increase accuracy, further technical work by AI talent with high-level expertise will likely be needed.

Even if such a solution were to be deployed in a localized setting, data availability bottlenecks would need to be overcome to ensure that the benefits could be realized on a broader scale. Alternatively, emerging techniques such as transfer learning and new model architectures could enable solutions to be derived from small data sets.

Data sets containing images of skin cancer and other conditions are not centralized and will need to be collected from many different healthcare providers. Potential solutions include consolidating provider data where possible on a platform and augmenting the database using physician-provided skin photos and diagnostic information. Another option would be to collaborate with large health systems including single country payors. In Finland, in an example illustrating the possibility of sharing healthcare data, the FinnGen genomic research partnership was able to create a system to gather and share blood sample data through biobanks. The University of Helsinki led the project, working with the Helsinki University Central Hospital and seven pharmaceutical companies. All Finns can participate in the research by giving consent to the biobanks to use their sample. Digital healthcare data from national registries were also provided to support the FinnGen research consortium.[29]

"Last mile" implementation issues could also be a bottleneck. The most prominent roadblock to deployment is the question of which party would be liable for misdiagnoses: the solution provider, the healthcare provider, or the insurer? This is not dissimilar to the unresolved accountability questions about self-driving cars. Once diagnosis through a mobile application is conducted, healthcare access, explanations, and coverage are needed. Specialized care is required to act on the diagnosis from a mobile application. This can be a challenge in many cases, including for rural communities. While smartphone penetration is rising globally, it remains below 50 percent in many developing regions (Exhibit 9).

Risks involved with mobile cancer diagnosis will need to be mitigated. Privacy violation is one such risk. Patient consent is required to leverage photos to be used for research purposes and diagnosis. Some countries do not allow healthcare data to leave the country (the data must be stored on servers within the countries' borders). A solution may be to build a global diagnostic model with anonymized databases, including country-specific databases when regulations apply, and require patient consent for doctors and others deploying the AI model. For example, 23andme users opt in for a service to allow for data collection. This requires explanation of the implications of opting in and adequate administrators to evaluate adherence to rules.[30]

Beyond privacy issues, there are also risks associated with ensuring safety. Legal repercussions could also result in case of misdiagnosis, though the question of who would be liable remains unresolved, as discussed above. A solution could be to require "human in the loop" intervention, for example a physician to validate

[28] Some commercial applications for this technology are also underway. See for example, Jonah Comstock, "SkinVision gets $7.6M to continue expanding skin cancer app," *Mobihealthnews*, July 30, 2018, www.mobihealthnews.com/content/skinvision-gets-76m-continue-expanding-skin-cancer-app.

[29] Mari Kaunisto, *FinnGen taps into a unique gene pool to find the next breakthroughs in disease prevention, diagnosis and treatment*, University of Helsinki FinnGen press release, December 19, 2017, helsinki.fi/en/news/health/finngen-taps-into-a-unique-gene-pool-to-find-the-next-breakthroughs-in-disease-prevention-diagnosis-and-treatment.

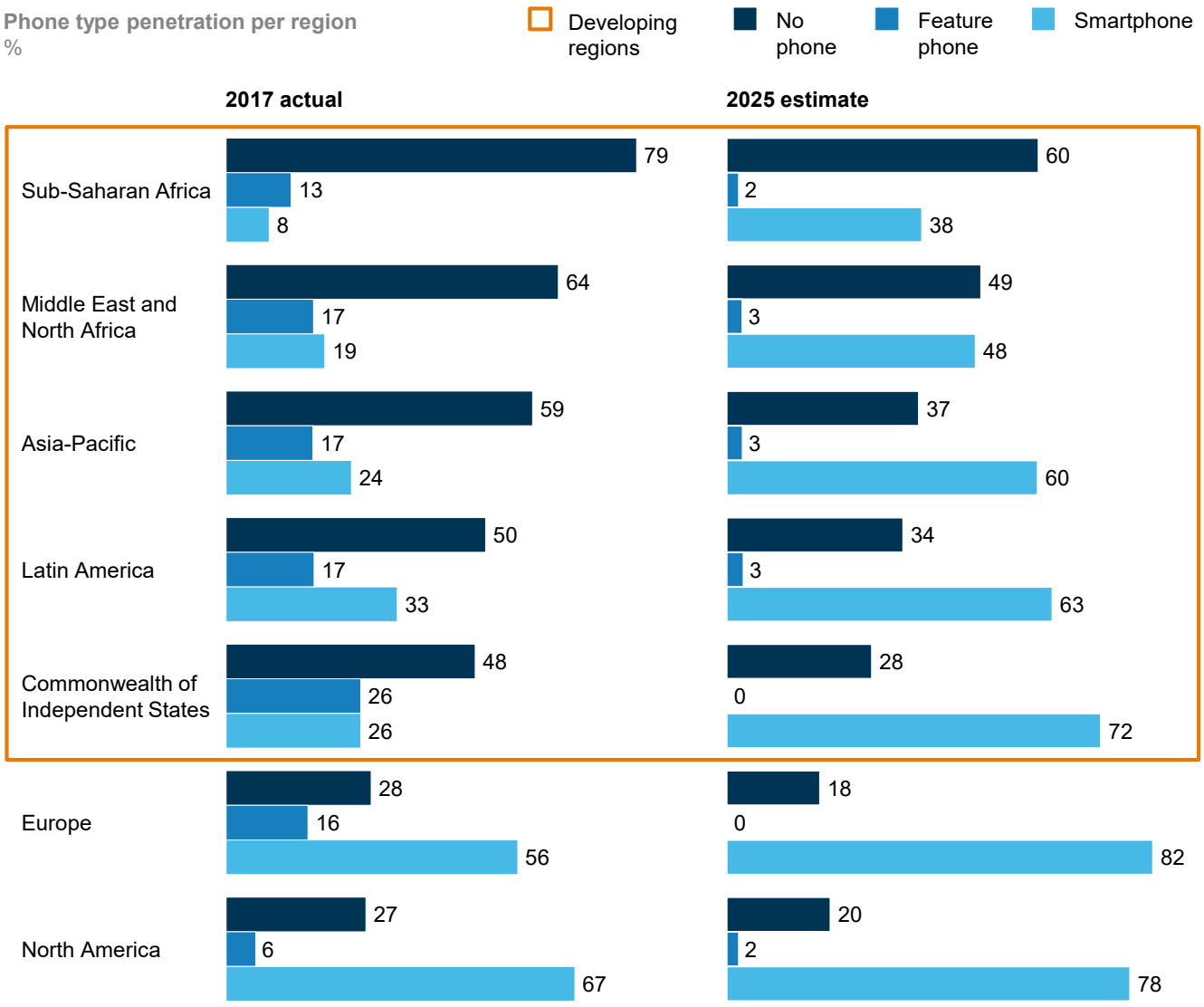
[30] *Research participation and consent*, 23andme, customercare.23andme.com/hc/en-us/articles/212195708-Research-Participation-and-Consent.

a diagnosis made by an AI solution.[31] Rigorous testing would also be required to ensure that diagnoses are accurate even if the context changes, for example, different skin tones or lesions in atypical parts of the body. Training data from people with different skin pigmentation will be required to enable effective and inclusive diagnosis.

[31] One team of researchers using machine learning to diagnose skin cancer discovered that their system reported a higher probability of skin cancer when the image included a ruler, because the training data included images of skin cancer that were typically accompanied by a ruler to illustrate the size of the lesion. Jan Bowers, "Should dermatologists fear machine learning?" *Dermatology World*, February 2018, digital. ipcprintservices.com/publication/?i=469994&article_id=2989257.

Exhibit 9

Smartphone penetration is rising but still low in emerging regions, where most visually impaired and unbanked or financially vulnerable people live.



NOTE: Feature phone figures assume equivalent to 2G penetration, while smartphone figures assume penetration of phones that use 3G or beyond. Figures may not sum to 100% because of rounding.

SOURCE: *The mobile economy 2018*, GSMA, 2018; McKinsey Global Institute analysis

3. Using AI models to improve financial inclusion in emerging economies

About 1.7 billion people worldwide are "unbanked," that is, they do not have an account at a financial institution or through a mobile money provider.[32] About half of them are from the poorest households in the global economy. Companies in some countries can also have difficulties in accessing finance; according to the World Bank, 20 percent of small and medium-sized enterprises in Africa cite finance as the biggest obstacle to maintaining their business.[33]

AI can help address core difficulties holding back financial inclusion: difficulties in verifying identities and in a lack of traditional data for underwriting services to vulnerable populations. AI-based financial inclusion products, which are already being deployed, bypass the need for a traditional credit score through analyzing digital footprints. Many fintech startups are entering the alternative credit-scoring space with AI-enhanced solutions, particularly in countries such as Bangladesh and Pakistan where the populations are large and significant portions are unbanked (Exhibit 10).

Companies such as CreditVidya, ZestFinance, and Lenddo capture alternative data by device, browser, and social media fingerprinting to generate a predictive model of creditworthiness. M-Shwari banking, which leverages the M-Pesa mobile money system in Kenya, incorporates telecommunications history in its assessment of credit risk. One in five adults in Kenya is currently an active user of the service, and M-Shwari is regarded as one of the most successful solutions for financial inclusion. Behind its SMS and internet-based interface, predictive algorithms leverage several AI capabilities to analyze social and telecom data and assess creditworthiness. The information is then processed in minutes and produces a credit score, which determines the size of the loan allowed.

A range of capabilities can be leveraged for these products, including natural language processing, structured deep learning, and person identification of social and telecom data. Long short-term memory recurrent neural networks can be trained to recognize an individual's credit risk. Image-processing capabilities can be used as an additional layer of verification to confirm an individual's identity. Structured and unstructured data from sources including social media, browsing history, telecom, and know-your-customer data can be used to train AI models. Solutions are likely to start with external data such as longevity as a telecom customer, and the model is then augmented against a client's actual product borrowing performance.

Successful implementation will require integration of multiple data sources

Key challenges to overcome include integrating multiple data sources, given different methods of storing information. The outcomes must be tested rigorously and explainable where necessary, given that an AI model is analyzing personal data to sort people, assess the credit risk of customers, and potentially reject some. Representative positive and negative data need to be collected to help reduce unwanted biases. Identifying the levers considered most strongly by the model when determining a credit score could help, as could providing information on the model decision-making process, particularly for rejected users.

"Last mile" implementation is also a potential bottleneck, because many adults in emerging economies who are unbanked also do not have mobile phones or internet access. This prevents the creation of a digital footprint, including telecommunications and online social history, which are the data needed for this form of alternative credit scoring. Overcoming such bottlenecks may require building partnerships with NGOs to provide funding for basic technology access.

Other implementation challenges may depend on the willingness of financial institutions to provide a mobile money infrastructure and support the use of alternate credit ratings, and on providing financial education and transparency to customers. For example, they will need to understand the implications of various actions on assessments of their creditworthiness.

In terms of risk—as with other solutions that draw inputs, including social media data and purchase histories, from highly personal information—data privacy and security are essential.

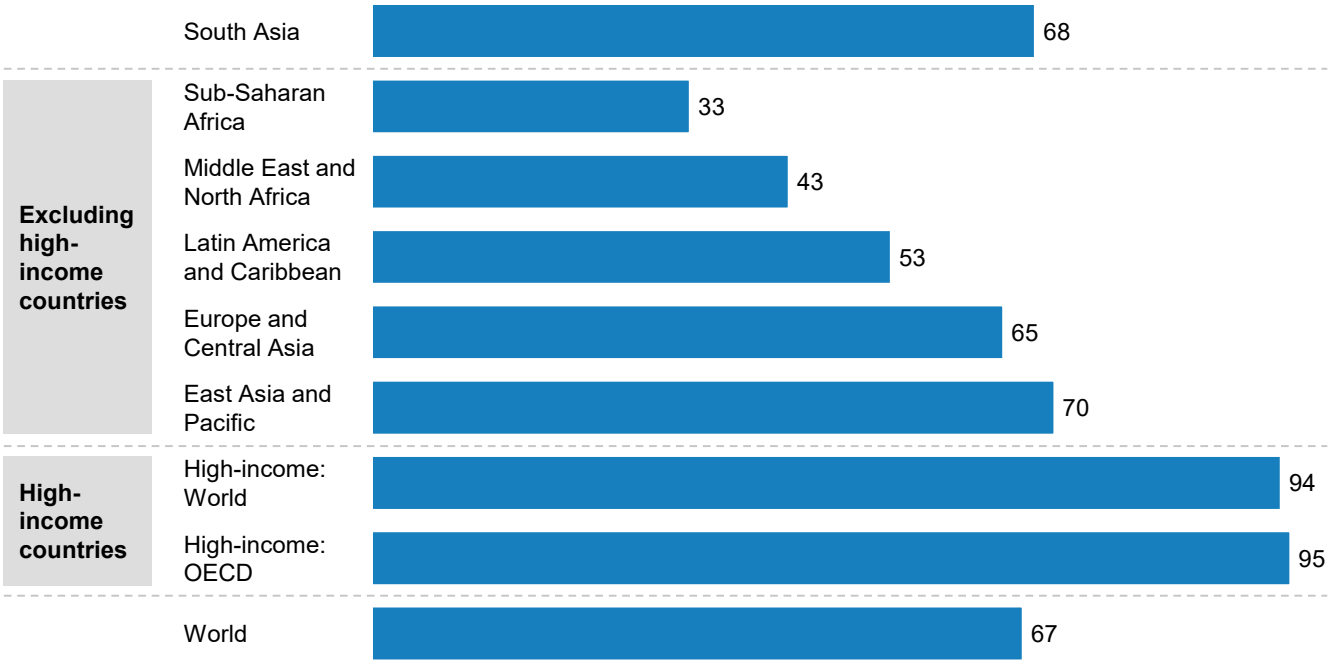
[32] For a detailed discussion of financial inclusion, see *Digital finance for all: Powering inclusive growth in emerging economies*, McKinsey Global Institute, September 2016, and the Global Financial Inclusion database, World Bank. globalfindex.worldbank.org.

[33] *Access to finance for small and medium enterprises in Africa*, World Bank, 2016, acetforafrica.org/acet/wp-content/uploads/publications/2016/03/Access-to-Finance-for-SMEs-Paper.pdf.

Exhibit 10

Financial inclusion is high in advanced economies but millions of people in emerging economies are unbanked.

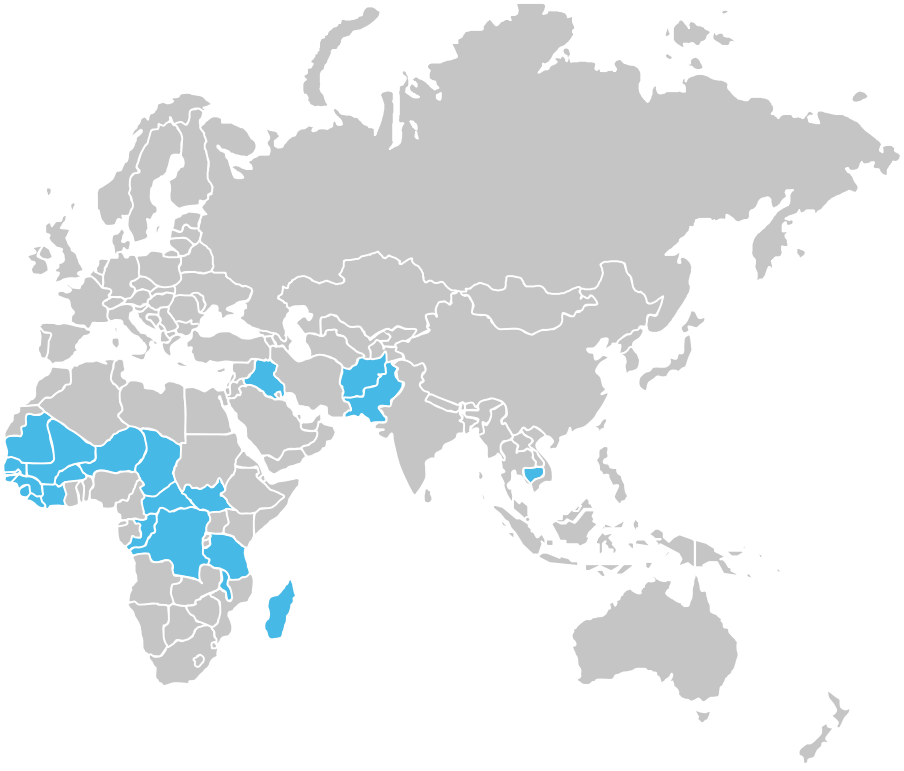
Estimated adult population with an account at a bank or other financial institution in 2017[1]
%



In 21 countries, fewer than 20% of adults had an account at a bank or other financial institution in 2017

Countries include

- Afghanistan
- Burkina Faso
- Cambodia
- Central African Republic
- Chad
- Democratic Republic of Congo
- Republic of the Congo
- Ivory Coast
- Guinea
- Iraq
- Liberia
- Madagascar
- Malawi
- Mali
- Mauritania
- Niger
- Pakistan
- Senegal
- Sierra Leone
- South Sudan
- Tanzania



1 Figures are based on percentage of respondents to a World Bank survey who indicated they have an account at a bank or other financial institution.

SOURCE: The World Bank Group Global Findex Database, globalfindex.worldbank.org; McKinsey Global Institute analysis

4. Using AI-powered drones to spot and stop poachers

The United Nations estimates that illegal wildlife trade worldwide could be worth $8 billion to $10 billion annually.[34] The value of the ivory trade alone contributes about $1 billion.[35]

For now, foot patrols and drone-based surveillance have not been effective at preventing poaching, and these efforts are labor-intensive and under resourced. One AI-based solution has already been built and tested, and it had some initial success in combating poachers. The SPOT system, built by researchers from the University of Southern California's Center for Artificial Intelligence in Society and piloted by the organization Air Shepherd, automates the process of detecting poachers in infrared video feeds, freeing park rangers for other tasks and increasing the reliability of surveillance (see illustration, "How AI can be deployed to catch wildlife poachers").[36]

The solution uses image classification and object detection to find animals and poachers on infrared video captured by a drone at night. A convolutional neural network model is trained to recognize both poachers and animals despite their small size in the video feed. The SPOT model does not need to be customized by the drone user and can be used in the field immediately, although a trained drone operator is still required.

Air Shepherd has reported some success in South Africa and plans for wider rollout in Botswana. In one area where as many as 19 rhinos were killed each month, there were no deaths for at least six months after the program was deployed.

With more development, SPOT and similar solutions could guide drones autonomously, adjusting flight routes to track poachers and removing the need for highly trained pilots and systems operators in the more than 300 wildlife parks in the world.

Scaling the AI solution for use in multiple parks will be needed to improve efficiency and reduce costs

Since each park has different flora, fauna, weather, and other conditions, data acquisition may need to be customized to optimize performance in any new park. To avoid having to generate a large amount of training data at the outset, the program could potentially be scaled to first target parks like those in the pilot programs, in order to minimize the labor required for additional training.

Real-time processing of data requires access to GPU-powered or other systems either in the cloud or on local computers. A local machine requires a significantly larger initial investment, while the cloud needs a reliable internet connection. Sharing a public cloud instance or a GPU-powered local machine between multiple parks would minimize the upfront investment as well as operating costs.

In general, hiring AI talent who can build and train this model could be challenging. Not all organizations looking to do something similar would have access to the high-level AI expertise that could develop the model, and even if they do, they may not have long-term support to refresh (or customize), troubleshoot, and improve on the model over time.

Implementation costs for state-of-the-art drones and infrared cameras could be significant for resource-constrained parks, although building partnerships with NGOs and governments to provide access to the required funds to purchase and maintain the equipment could be a cost-sharing solution.

Implementation talent (non-AI) is also a challenge: while SPOT will eventually pilot the drone, the aircraft will still require trained personnel for launching, maintaining, and troubleshooting of equipment, including the infrared camera. The limited supply and cost of employing these professionals can limit scalability. Scaling the program could involve parks sharing drone operators, driving down costs. Drone operators can also train park personnel to use the drones. Once they are proficient, park personnel would need to contact the professional operators only for specialized repairs.

---

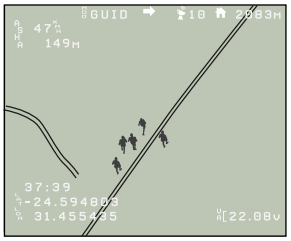
[34] *Wildlife crime worth USD 8-10 billion annually, ranking it alongside human trafficking, arms and drug dealing in terms of profits*, United Nations Office on Drugs and Crime, May 13, 2014, unodc.org/unodc/en/frontpage/2014/May/wildlife-crime-worth-8-10-billion-annually.html.

[35] Jason Bellini, "Price of ivory? 100 elephants killed per day," *Wall Street Journal*, March 19, 2015, blogs.wsj.com/briefly/2015/03/19/price-of-ivory-100-elephants-killed-per-day-the-short-answer/.

[36] Elizabeth Bondi et al., *SPOT poachers in action: Augmenting conservation drones with automatic detection in near real time*, 32nd AAAI Conference on Artificial Intelligence, April 27, 2018, teamcore.usc.edu/papers/2018/spot-camera-ready.pdf.

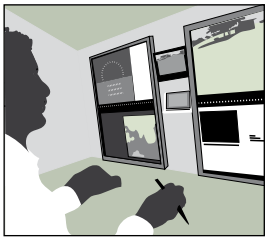
How AI can be deployed to catch wildlife poachers

Six steps from offline training of AI model to online detection

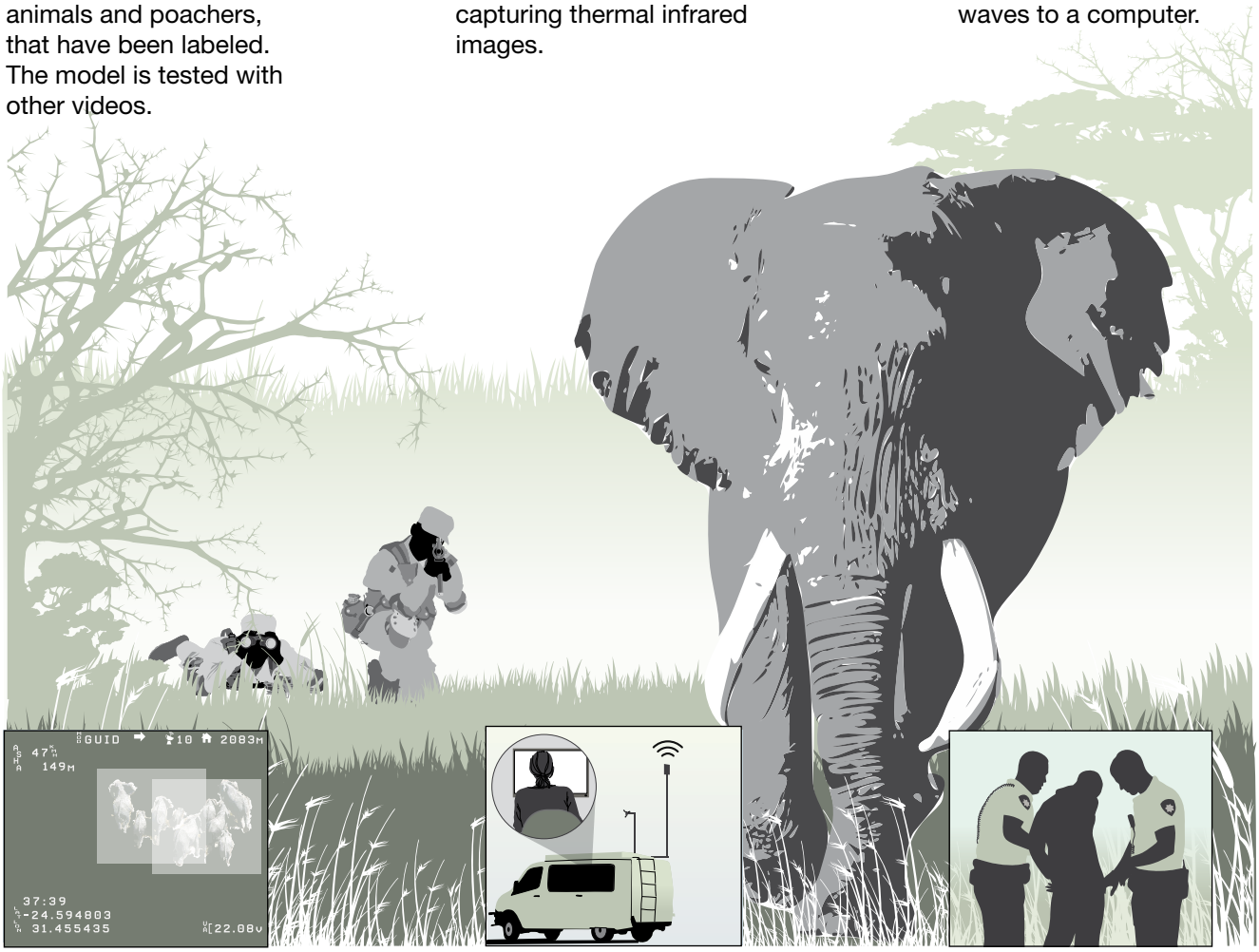


1. Offline training
A neural network is trained on 70 videos, containing animals and poachers, that have been labeled. The model is tested with other videos.

2. Drone deployment
Drones are flown over wildlife sanctuaries, capturing thermal infrared images.



3. User interface
Video and still images are transmitted via radio waves to a computer.



4. Pre-processing
The infrared images may need to be converted to "white-hot" format, where warm objects are lighter against a dark background.

5. Detection
The video is processed in batches and sent to the cloud for analysis. Each image is treated as an input into the neural network.

6. Output
The neural network outputs annotations that are overlaid on top of the original image. This enables identification of the poachers' whereabouts.

SOURCE: Elizabeth Bondi et al., *SPOT poachers in action: Augmenting conservation drones with automatic detection in near real time*, 32nd AAAI Conference on Artificial Intelligence, April 27, 2018; McKinsey Global Institute analysis

5. In Flint, Michigan, detecting water service lines containing lead

In the water crisis in Flint, Michigan, lead contamination of water from pipes resulted in a lead concentration in the water supply that was 300 times the U.S. Environmental Protection Agency limit.[37] About 9,000 children aged six or younger were exposed to the risk of permanent harm to brain development, impaired learning abilities, and behavioral disorders.[38]

The primary contributor to lead in the water system is thought to be water service lines, which connect the city water supply to homes. However, finding and replacing water service lines in Flint requires excavation, because of a lack of records about pipe materials. This is where AI and advanced analytics can come in. The University of Michigan has developed a model called ActiveRemediation that can predict with 98 percent accuracy whether a water service line is lead. Deployment of the predictive model in Flint reduced unnecessary replacement excavations from 18.8 percent to 2 percent.[39]

While the other use cases in this section deploy deep learning, this example combines an instance of machine learning—often considered an AI technique—and Bayesian analysis to make pipe material predictions. It uses a combination of city parcel data, census data, and incomplete records of water service lines to predict the homes most likely to have lead pipes (see illustration, "How the lead pipe detection model in Flint, Michigan, works"). In addition to the application in Flint, this approach could be relevant in other water pipe replacement efforts, although local data would be required. As the case illustrates, a broad view of analytical and AI techniques that could be applicable to any given problem can be more appropriate in some cases than using only deep learning.

Data availability is a key to this solution and will need to be scaled

A stringent and systematic data-gathering process by the city is key. Digitization of such data is also a prerequisite for AI deployment. One of the cost factors in Flint was the lack of digital records.

Implementation was relatively straightforward, but the challenge will be to scale this type of solution. The pipe material model was developed specifically for Flint with the data available in Flint. That means other data sources will be required for application to other areas.

[37] Christopher Ingraham, "This is how toxic Flint's water really is," *Washington Post*, January 15, 2016, washingtonpost.com/news/wonk/wp/2016/01/15/this-is-how-toxic-flints-water-really-is.

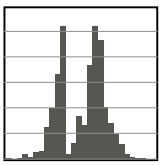
[38] The University of Michigan's ActiveRemediation model for detecting lead pipes is described in detail in Jacob Abernethy et al., *ActiveRemediation: The search for lead pipes in Flint, Michigan*, Cornell University Library, August 17, 2018. https://arxiv.org/abs/1806.10692. See also *Lead poisoning and health*, World Health Organization, August 23, 2018, www.who.int/news-room/fact-sheets/detail/lead-poisoning-and-health.

[39] Ibid. Jacob Abernethy et al., Cornell University Library, August 17, 2018.

How the lead pipe detection model in Flint, Michigan, works

Six steps from offline training of AI model to online detection

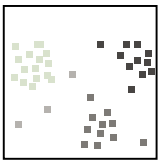
Data collection



1. City of Flint provided team with a dataset describing each of the 55,893 parcels in the city. It includes attributes of each home, such as property owner, address, value, and building characteristics.

2. Digitization of city records of service lines using OCR on data, such as hand-annotated maps containing markings for each parcel and incomplete records of materials used for each home's service line.

Model building



3. Machine learning model predicts probability that the portion of a service line attached to a particular home is hazardous based on known features, such as home value.

4. A Bayesian model is used to fine-tune predictions to account for relationships between parcels and precincts.

5. Algorithm determines which homes to observe next, by efficiently allocating resources.

6. Determination of final list of homes to be visited by the replacement crew, based on the probability that they have lead pipes.

Blocks of houses in the Grand Traverse district in Flint, Michigan. Percentages on each house denote the machine learning–derived probability of lead water pipes.

SOURCE: Jacob Abernethy et al., *ActiveRemediation: The search for lead pipes in Flint, Michigan*, Cornell University Library, August 17, 2018; McKinsey Global Institute analysis

McKinsey&Company

6. Analyzing satellite data to help cope with the aftermath of natural disasters

Natural disasters kill more than 50,000 people and displace tens of millions more each year. The total economic damage amounts to more than $100 billion annually. Populations in less developed countries are hardest hit: in countries with a medium or low human development index, up to six times as many people can be affected by natural disasters compared to more affluent countries, and the economic damage can be four times greater.[40]

To coordinate and prioritize emergency response, governments and first responders must have an accurate and complete view of disaster zones. Frequent and broad area satellite imagery enables new AI-based systems to quickly and accurately detect infrastructure changes that may affect evacuation and response. AI can assist in improving relief efforts and emergency preparedness with greater accuracy and on a much larger scale than human workers. For example, following the passage of Hurricane Harvey in 2017, a collaboration between Planet Labs, provider of satellite imagery, and CrowdAI provided an immediate view of the greater Houston area and was able to detect road outages due to flooding and quantify infrastructure damage.

The solution leverages computer vision capabilities—specifically, object detection—to determine which portions of satellite imagery belong to the target feature. In the case of detecting road outages, the model was able to identify all roads in a satellite image. Other assessments of critical infrastructure damage required identification of objects such as building outlines. The solution was also able to identify the presence or absence of water in broad areas affected by floods on false color satellite imagery using various analytic methods. Combined, the resulting mapping provides an accurate view of usable roads and undamaged buildings, updated daily as the satellite imagery is refreshed.[41]

Satellite data can be used for many other solutions ranging from short-term weather forecasts to tracking deforestation. Indeed, satellite data can power AI applications across all ten of our domains. This is especially the case for infrastructure and economic empowerment, where advances in satellite imagery resolution enable visualization and monitoring of more granular portions of Earth, such as small agricultural fields in developing nations or clusters of buildings in urban areas. Satellite imagery can also provide perspective into infrastructure projects in remote, inaccessible locations where on-foot monitoring is often infeasible due to the arduous, time-consuming nature of collecting information on the ground. The Indian government has employed satellite-based monitoring systems to understand the status of projects on a timely basis and detect where fund allocation is not effective.[42]

Use cases that pair satellite data with other data include augmenting remote sensing data from satellites with prices to create a forecast of agricultural production. Satellite data could also be used to measure economic activity by detecting car density, construction rates, and electricity consumption through nighttime illumination. The following illustration is a photograph showing how object detection software was applied to satellite imagery to detect flooded roads after Hurricane Harvey in 2017. The object detection software is trained to detect and map roads in imagery. During a flood, the software can rapidly identify which roads are no longer distinguishable, and therefore impassable, due to flood waters. In the photograph, flooded roads are marked in red, while nonflooded ones are in orange.

Access to satellite data and rigorous testing are essential for AI use for disasters

To use satellite data in disaster scenarios, access is the key challenge to overcome. Most satellite data are held by institutions that often do not make their data freely available, and access to the data may be prohibitively expensive for NGOs. However, most space agencies and space system operators—in addition to some nongovernmental satellite data owners—belong to the International Charter on Space and Major Disasters. Through that worldwide collaboration, satellite data are made available at no cost for purposes of disaster management.

In terms of risk, as with other solutions that provide critical information, the need for rigorous testing to ensure accuracy is essential. If the solution incorrectly indicates that a road is clear of flooding and directs thousands of people toward it, there could be significant consequences, such as increased risk of harm and reduced evacuation speed.

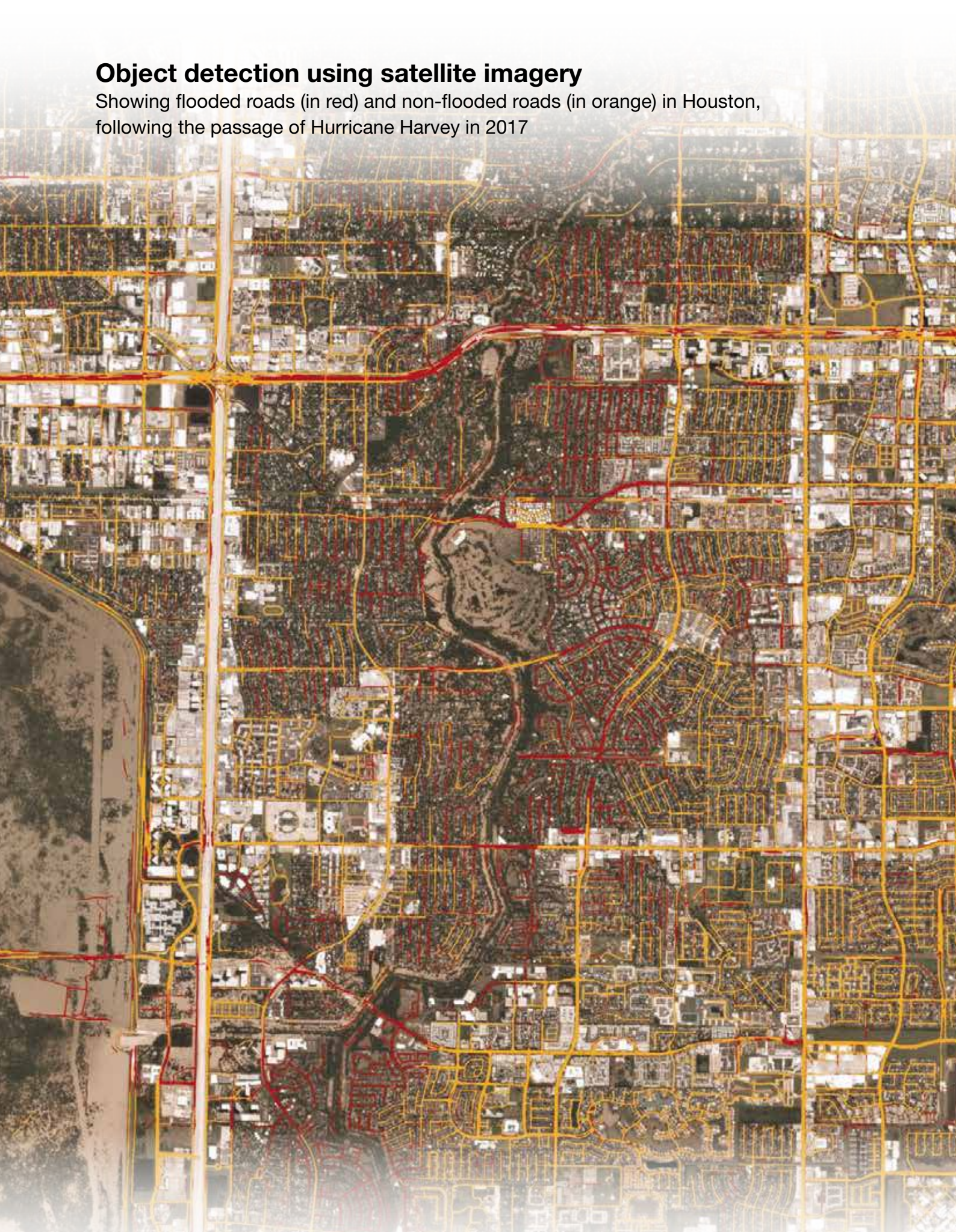
[40] More affluent countries are defined as those with either a very high or high human development index. Hannah Ritchie and Max Roser, "Natural catastrophes," *Ourworldindata.org*, 2018, ourworldindata.org/natural-catastrophes.

[41] "Anatomy of a catastrophe: Using imagery to assess Harvey's impact on Houston," Planet.org, 2018, planet.com/insights/anatomy-of-a-catastrophe/.

[42] Santosh Subramanian, "Infrastructure monitoring: A case study," SkyMap Global, June 19, 2018, skymapglobal.com/infrastructure-monitoring-with-satellite-imagery-a-case-study/.

Object detection using satellite imagery

Showing flooded roads (in red) and non-flooded roads (in orange) in Houston, following the passage of Hurricane Harvey in 2017



SOURCE: PlanetLabs; McKinsey Global Institute analysis

4. BOTTLENECKS TO OVERCOME

While the social impact of adding AI to the mix of solutions targeting the world's most pressing challenges is potentially very large, some AI-specific bottlenecks will need to be overcome if even some of that potential is to be realized. In all, based on interviews with social domain experts and AI researchers, we identified 18 potential bottlenecks that could stand in the way of successful AI deployments for social good. Having identified the 18, we then tested them on use cases in our library. Exhibit 11 shows these bottlenecks grouped in four categories of criticality.

Exhibit 11

Bottlenecks limiting the use of AI for societal good can be grouped into four categories.

Critical for most domains

- Data accessibility
- Data quality
- High-level AI expertise availability
- High-level AI expertise accessibility
- Regulatory limitations
- Organization deployment efficacy

Critical for select cases[1]

- Data volume
- Data labeling
- AI practitioner talent availability
- AI practitioner talent accessibility
- Access to computing capacity

Contextual challenges

- Data availability
- Data integration
- Access to technology
- Organization receptiveness
- Public resistance

Potential bottleneck

- Availability of organizations that can scale AI deployment
- Access to software libraries and other tools

1 Bottlenecks that are critical for some domains as a whole or for individual use cases within those domains.
NOTE: This list of bottlenecks was derived from interviews with social domain experts and AI researchers and tested against our use case library.

SOURCE: McKinsey Global Institute analysis

The most significant bottlenecks we identified, and which we describe in detail in this chapter are data accessibility, a shortage of talent to develop AI solutions, and "last mile" implementation challenges.[43]

The availability of computing infrastructure on which to train AI models is no longer as significant a barrier as it once was. Competition among vendors has greatly reduced the cost of cloud-based computing capacity, including graphics processing units (GPUs), and increased access to affordable computing capacity.[44] Cloud-based computation as a service is now widely accessible and requires relatively small investment on a pay-as-you-go basis. For example, fast.ai has shown that, using its framework and a cloud computing instance, a model can be trained on the ImageNet corpus of images to create a model with a 93 percent accuracy rate to identify objects in images for only $25.[45]

[43] Other AI research we have conducted suggests that AI's application could be unequal, with a divide opening up between countries and companies. See Jacques Bughin and Nicolas van Zeebroeck, "The promise and pitfalls of AI," *Project Syndicate*, September 6, 2018, https://www.project-syndicate.org/commentary/artificial-intelligence-digital-divide-widens-inequality-by-jacques-bughin-and-nicolas-van-zeebroeck-2018-09.

[44] Access to GPU capacity can nonetheless be tight in the short term.

[45] Jeremy Howard, *Training Imagenet in 3 hours for $25; and CIFAR10 for $0.26*, fast.ai, May 2, 2018, fast.ai/2018/04/30/dawnbench-fastai/.

Similarly, advances in open-source AI libraries have substantially simplified many of the tasks previously requiring significant programming skill and knowledge of AI algorithms. Software libraries such as PyTorch, TensorFlow, and fast.ai are available to a global public on an open-source basis, and the ease with which they can be deployed and used continues to improve over time.

DATA NEEDED FOR SOCIAL IMPACT USES MAY NOT BE EASILY ACCESSIBLE FOR NGOS AND OTHERS IN THE SOCIAL SECTOR

Data barriers we identified are primarily tied to issues of the accessibility, quality, and quantity of data. Data volume issues—essentially, the difficulty of obtaining sufficiently large amounts of rich data, including input from video, image, and text with which to train algorithms—remain an obstacle. However, as the AI field continues to advance, and more models are pretrained with large amounts of data in various domains, the incremental amount of data required to solve individual problems can often be reduced. For example, transfer learning, in which an AI model is trained to accomplish a certain task and then applies that learning to a similar but distinct activity, will reduce the requirement for massive training data volume for each individual activity.

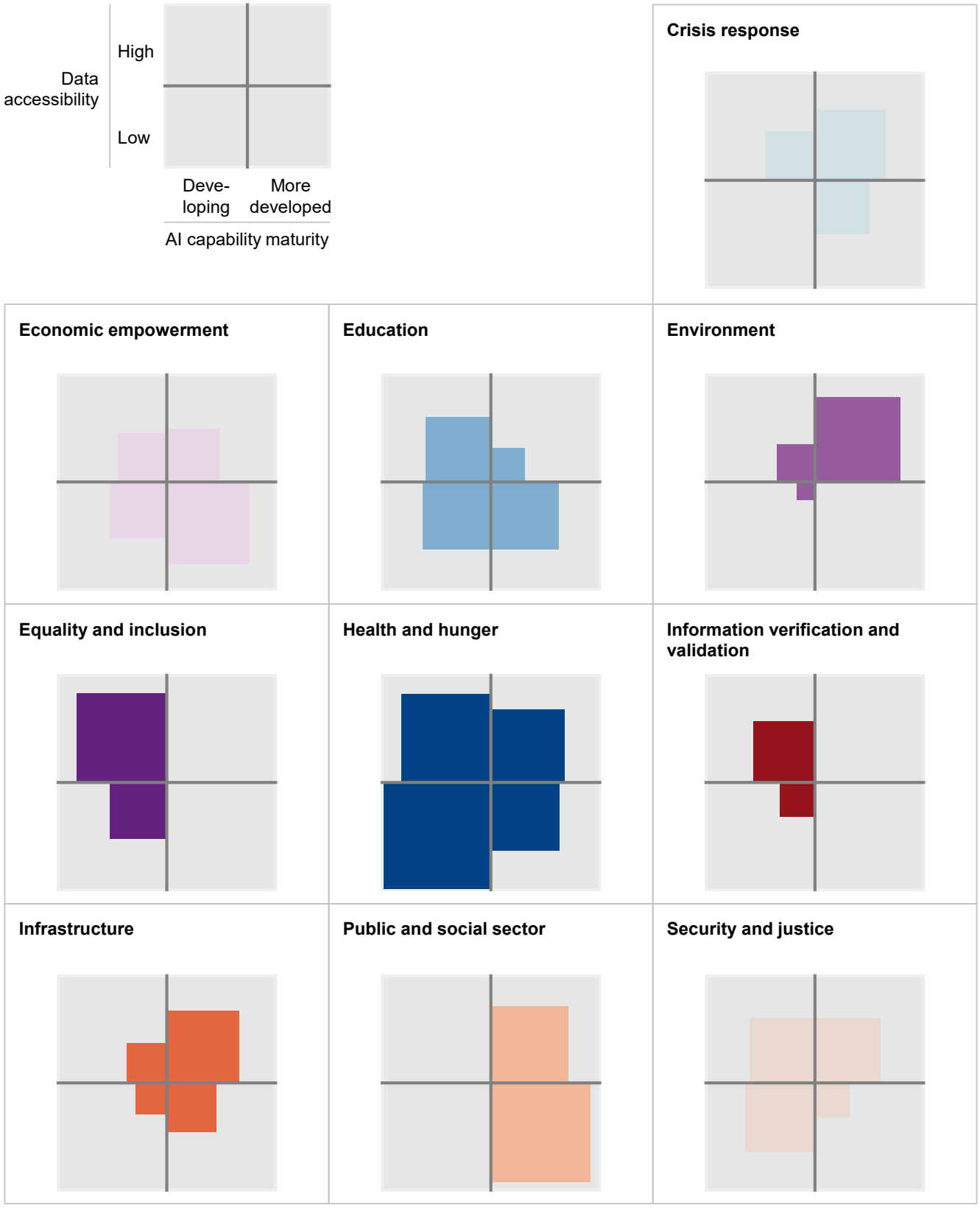
Data accessibility remains a significant challenge. To resolve it will require a willingness by both private- and public-sector organizations to share their data. Much of the data that are essential or useful for social good applications are in private hands or in public institutions that might not be willing to share their data. Organizations with large data sets include telecommunications and satellite companies, social media platforms, financial institutions (for details such as credit history), hospitals, doctors and other health providers (medical information), and governments (including private individuals' tax information). Obtaining access to these types of data sets by social entrepreneurs and NGOs can be difficult because of regulations on data use, privacy and other concerns around risks, and bureaucratic inertia. The data may have business value and be commercially available for purchase. Given the challenges of distinguishing between social use and commercial use, the price may be too high for NGOs and others wanting to deploy the data for societal benefit. Owners of the data may also be unwilling to dedicate the time and resources needed to clean and share data.

Indeed, data quality can be its own challenge. If the data used to build and run accurate and fair AI models are not representative or of sufficiently high quality—for example, if some data are partially missing, outdated, or contain errors—this can be a serious risk, as we discuss in more detail in Chapter 5. Exhibit 12 shows how our use cases map to data accessibility and AI capability maturity across domains. For example, the crisis response and public and social sector management domains require mostly mature capabilities, while solutions for equality and inclusion are still developing.

Exhibit 12

A mapping of use cases to data accessibility and AI capability maturity shows that the impact is highly varied.

Sum of usage frequency score of use cases mapped to level of data accessibility and capability maturity barriers[1]



1 Capability maturity refers to the combination of the level of talent needed to build a solution and the level of maturity of the capabilities required.
NOTE: Our library of about 160 use cases with societal impact is evolving and this chart should not be read as a comprehensive gauge of the potential application of AI or analytics capabilities.

SOURCE: McKinsey Global Institute analysis

THE EXPERT AI TALENT NEEDED TO DEVELOP AND TRAIN AI MODELS IS IN SHORT SUPPLY IN THE SOCIAL SECTOR

While just over half the use cases in our library can leverage solutions that can be created by talent with relatively lower levels of AI experience, the remaining use cases have added complexity due to a combination of factors, depending on the specific case. These need high-level AI expertise, that is, people who may have PhDs or considerable experience with the technologies—and who are in short supply.

For the use cases requiring less AI expertise, the solution builders needed are data scientists or software developers with AI experience but not necessarily high-level expertise. This is because some AI capabilities are less technically complex to deploy. Most of these use cases rely on single modes of data input.

Problem complexity increases significantly where use cases rely on several AI capabilities to work together cohesively and require multiple different data-type inputs. Making progress in developing solutions to these cases will thus require high-level talent, for which demand far outstrips supply and competition is fierce.[46] Reflecting this supply-demand imbalance, compensation for AI practitioners with high-level expertise is very high in the commercial sector, straining the ability of social-sector organizations to recruit. Some areas of research tend to attract talent more than others; for example, ambitious "moonshot" challenges may be more attractive to some experts than practical AI applications.

"LAST MILE" IMPLEMENTATION CHALLENGES ARE ALSO A SIGNIFICANT BOTTLENECK FOR AI DEPLOYMENT FOR SOCIAL GOOD

Even when high-level AI expertise is not required, NGOs and other social-sector organizations can still face technical problems in deploying and sustaining AI models for which they will need continued access to some level of AI-related skills. The talent required might include engineers who can maintain or improve on the models, data scientists who can extract meaningful output from AI models, and so on. Failed handoffs will occur when solutions providers only set up the solution and then disappear without ensuring that a sustainable plan is in place.

One example cited by experts we interviewed concerned an AI-powered research tool built for a federal agency. The agency did not understand the technical documentation describing how to install and run the tool, and the tool became unused "shelfware" once the agency's contract with the private research group that devised the solution expired. To avoid such problems, the organization deploying the AI solution could either ensure that it has the capability to maintain and operate a tool in the long term or contract for technical support, including updating and maintaining the model.

Organizations may also have difficulty interpreting AI model results. Even if a model achieves a desired level of accuracy on test data, new or unanticipated failure cases can often emerge in real-life scenarios. Without an understanding of how the solution works, which may require data scientists or "translators"—that is, people who can bridge the technical expertise of data engineers and data scientists with the operational expertise of frontline managers—the NGO or other implementing organization may be overly trusting of the model results, even though most AI models cannot perform accurately all the time. The models are often described as "brittle," that is, failing when inputs stray in specific ways from the data sets on which the models were trained.[47]

---

[46] Cade Metz, "Tech giants are paying huge salaries for scarce A.I. talent," *New York Times*, October 22, 2017.

[47] See Nicolaus Henke, Jordan Levine, and Paul McInerney, "You don't have to be a data scientist to fill this must-have analytics role," *Harvard Business Review*, February 5, 2018, hbr.org/2018/02/you-dont-have-to-be-a-data-scientist-to-fill-this-must-have-analytics-role.

Other “last mile” implementation challenges that are not technical in nature may seem more mundane but could be equally important and difficult to resolve. Change management of organizations, including adapting processes to be able to integrate AI-powered solutions, is one of the largest obstacles for both commercial and noncommercial deployment. Another potential hurdle is a lack of critical infrastructure. At the individual level, for example, finding an AI-powered solution that uses smartphones will be of little use to people who do not have them. Likewise, in areas without electricity, even simple devices that require recharging will not achieve the desired effect. Funding could help overcome such limitations but may need to be on a very large scale. For example, 90 percent of the 215 million visually impaired persons worldwide who could benefit from environment understanding software on smartphones live in developing countries, where smartphone penetration is low. Obtaining financial commitments for these investments, reductions in the costs of technology, or both will be needed to solve this challenge.

5. RISKS TO BE MANAGED

Risks associated with AI are becoming an increasingly important area of research, especially (but not exclusively) in the field of ethics applied to AI. AI's tools and techniques can be misused by authorities and others with access to them, and principles for their use will need to be established (see Box 2, "A growing body of research on the ethics of AI"). In the worst case, AI solutions can unintentionally harm the very people they are supposed to help.

Some AI-related risks spring from the way AI models are trained. For example, if data sets used to train algorithms are based on historical data that incorporate racial or gender bias (even unintentionally, resulting solely from sampling bias), the applications derived from the algorithms will perpetuate and may aggravate that bias.

In general, risks relating to AI for social good are quite similar to those for more commercial uses. One of the biggest risks is that AI's tools and techniques can be misused by authorities and others with access to them; malicious uses can harm individuals, organizations, and society at large.[48] AI can be used maliciously to threaten the physical and emotional safety of individuals, as well as their digital safety, financial security, and equity and fair treatment. For organizations, malicious use often implies reputation and legal compliance risks, although there may be fewer commercial risks than those that could potentially harm for-profit companies. For society, malicious uses of AI could threaten national security, economic stability, political stability, labor market stability, and infrastructure.

One general difference between the risks associated with commercial and noncommercial purposes concerns the effects of labor displacement on workers and the workforce. Much of the public debate around AI in the for-profit world focuses on the potential for labor displacement. An examination of our use case library suggests that this may be less of a risk in many of the applications of AI for social good. Indeed, as outlined in the earlier discussion of bottlenecks, AI's application for social good uses tends to be hampered by a shortage of personnel with the skills and technical know-how required. If AI were to increase access to health or education, for example, this could serve as a potential net boost to employment of doctors, nurses, or teachers.

An analysis of our use case library found four main categories of risk that are particularly relevant when leveraging AI solutions for social good, as we describe below. They are bias and fairness, privacy, safe use and security, and "explainability"—being able to identify the feature or data set that leads to a particular decision or prediction. The types of risk and their magnitude differ considerably from case to case. All domains carry some level of risk, but in general our analysis suggests that domains in which data are sensitive and predictions identify individuals—for example economic empowerment, education, equality, health, and security—face the highest magnitude of risk (Exhibit 13).

Inaccurate AI poses high risks in other domains, such as crisis response. For example, erroneous predictions of the location of missing persons could prove fatal. The scoring in the exhibit is based on perspectives from our interviews with domain experts and AI specialists. To come up with the risk profiles, we tagged individual use cases from low to high for each of the five categories: the risk of bias; the risk of privacy violation; the risk of unsafe use of the AI solution; the level of explainability that is required to reduce or mitigate risks; and considerations of the risk of negative impact to the workforce and workers.

[48] See Miles Brundage et al., *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*, Future of Humanity Institute, February 2018.

Box 2. A growing body of research on the ethics of AI

As AI capabilities develop and are increasingly deployed in both the commercial and noncommercial worlds, the ethics surrounding use of artificial intelligence have spurred a growing body of research.

Donors including technology company executives have stepped up funding for major scholarship programs in the past months, including at MIT, Harvard's Berkman Klein Center for Internet and Society, Stanford University, and the University of Toronto, to study the implications of AI, including how it will affect people's lives and serve humanity.[1]

Technical professional bodies and international bodies including the World Economic Forum and the UN's International Telecommunication Union have also focused attention on societal uses and potential abuses of AI.[2] The ITU has initiated an AI for Good conference that ties AI outcomes to the UN Sustainable Development Goals.

Several nonprofit organizations also do important work on ethical uses of AI, including the Partnership on AI, which brings together academics, researchers, civil society organizations, and companies building and utilizing AI technology to better understand AI's impacts.[3] And major technology companies including Microsoft and Google have articulated their philosophy and practices on AI; Google's first principle is "be socially beneficial."[4]

One recurring concern is the impact of AI on work, and whether it could create a gap even wider than the existing digital divide, not just between high- and low-income workers but also between countries.[5] While proposed solutions to these potential gaps differ, there is broad agreement that a substantial change in the way we educate our children is needed to prepare them for the future of work, and that large-scale retraining will be required for midcareer workers as skill requirements shift.[6]

Safety is a concern in social impact uses, just as it is in commercial use, for example with self-driving cars, since accidents can create ethical dilemmas in life-and-death situations. In healthcare-related uses, errors and decisions can systematically burden groups of people if solution design is flawed. An example of this is a diagnostic tool that may overall be better than a human doctor but that makes significantly more diagnosis errors when the patient is a member of an ethnic minority.

---

1 Steve Lohr, "M.I.T. plans college for artificial intelligence, backed by $1 billion," *New York Times*, October 15, 2018; Maria Di Mento, "Donors pour $583 million into artificial-intelligence programs and research," *Chronicle of Philanthropy*, October 15, 2018.

2 See, for example, *How to prevent discriminatory outcomes in machine learning*, World Economic Forum, March 12, 2018. The Institute for Electrical and Electronics Engineers has announced the approval of three standards projects inspired by the work of its Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. The new standards projects directly reflect the institute publication. See Kyarash Shahriari and Mana Shahriari, *IEEE standard review—Ethically aligned design: A vision for prioritizing human well-being with artificial intelligence and autonomous systems*, IEEE Canada International Humanitarian Technology Conference, Toronto, Canada, July 21–22, 2017.

3 McKinsey is a member of the Partnership on AI, partnershiponai.org/.

4 See *Microsoft AI principles*, microsoft.com/en-us/ai/our-approach-to-ai, and The Keyword, "AI at Google: Our principles," blog entry by Sundar Pichai, June 7, 2018, blog.google/technology/ai/ai-principles/.

5 Erik Brynjolfsson and Andrew McAfee, *The Second Machine Age*, New York: W. W. Norton, 2016.

6 See *Jobs lost, jobs gained: Workforce transitions in a time of automation*, McKinsey Global Institute, December 2017; *Skill shift: Automation and the future of the workforce*, McKinsey Global Institute, May 2018.

Box 2. A growing body of research on the ethics of AI (continued)

One of the most intense areas of research concerns bias. While algorithms and AI can improve on human decision making to reduce bias in some cases, the models can also end up reflecting intrinsic bias contained in the data sets used to train them.[7] For example, a recent report by MIT and Stanford researchers showed how some commercially available facial recognition systems perform poorly when applied to faces of women and people of color.[8] A study by Harvard's Berkman Klein Center shows how some popular AI use cases could both positively and negatively impact specific aspects of the Universal Declaration of Human Rights, depending on how they are implemented.[9]

Given the potential for negative impacts, some call for AI to be regulated by government. The benefits of regulation will need to be balanced against potential unintended consequences, including a dampening of innovation.[10]

Many governments have produced formal AI strategies to promote the use and development of AI, and most include specific sections on the development of ethics regulations.[11]

History shows that many new technologies have created misgivings, dating back at least to the 16th century with the invention of the stocking frame.[12] As has happened in the past, all stakeholders, from civil society to AI researchers to government, will need to collaborate to define what is—and is not—acceptable, if the positive benefits that the technologies offer are to become a reality.

---

7 One example of AI being less biased than humans is in bail decisions; one paper has found that algorithmic decision making in bail decisions could reduce jail populations by 42 percent with no increase in crime rates. Jon Kleinberg et al., *Human decisions and machine predictions*, NBER working paper number 23180, February 2017.

8 Joy Buolamwini and Timnit Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," *Proceedings of Machine Learning Research*, 2018, Volume 81, http://proceedings.mlr.press/v81/buolamwini18a.html.

9 Filippo Raso et al., *Artificial intelligence and human rights: Opportunities and risks*, Berkman Klein Center for Internet and Society at Harvard University, September 25, 2018, https://cyber.harvard.edu/sites/default/files/2018-09/2018-09_AIHumanRightsSmall.pdf.

10 Christopher Fonzone and Kate Heinzelman, "Should the government regulate artificial intelligence? It already is," The Hill, February 26, 2018, https://thehill.com/opinion/technology/375606-should-the-government-regulate-artificial-intelligence-it-already-is.

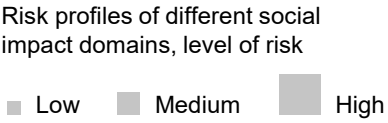
11 See for example, Cédric Villani, *For a meaningful artificial intelligence: Towards a French and European strategy*, March 2018, aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf; Will Hurd and Robin Kelly, *Rise of the machines: Artificial intelligence and its growing impact on U.S. policy*, US House of Representatives, Subcommittee on Information Technology, September 2018, oversight.house.gov/wp-content/uploads/2018/09/AI-White-Paper.pdf; Pan-Canadian artificial intelligence strategy, CIFAR, https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy; Tim Dutton, "An overview of national AI strategies," *Medium*, June 28, 2018, medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd.

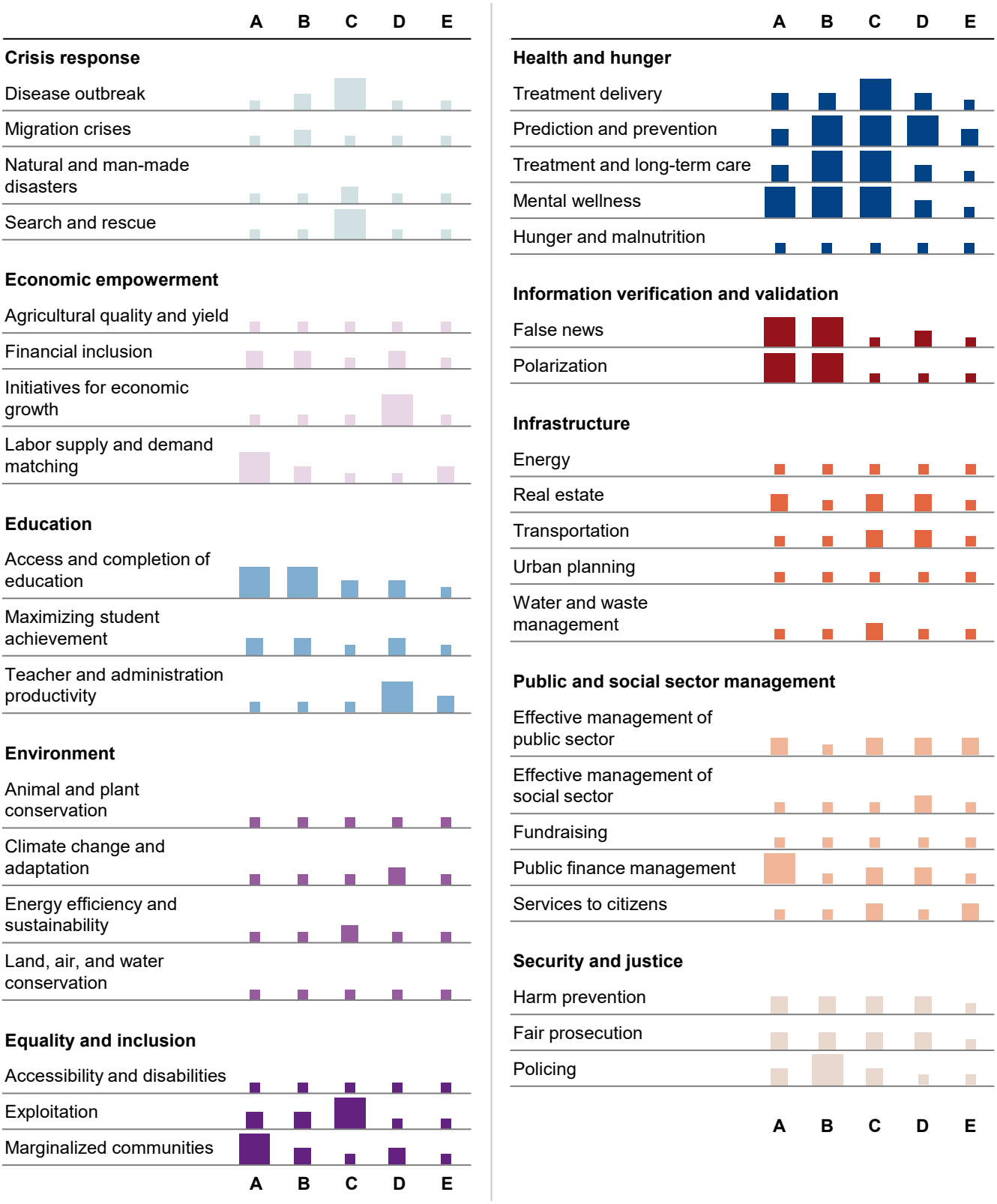
12 In 1589, England's Queen Elizabeth I refused to grant a patent to a stocking frame invented by William Lee because she was supposedly concerned about the effect on hand knitters. R. L Hills, "William Lee and his knitting machine," *Journal of the Textile Institute*, July 1989, Volume 80, Number 2.

Exhibit 13

Risk profiles of social impact domains differ, with some of the biggest potential risks found in the health and hunger domain.

A Bias **B** Privacy violation **C** Unsafe use
D Level of explainability required **E** Potential for workforce displacement





NOTE: The risk scoring in this exhibit is based on interviews with domain experts and AI specialists. We tagged individual use cases in our library from low to high in each domain and then aggregated the score to the issue type level. Our use case library continues to evolve and these findings should not be read as exhaustive.

SOURCE: McKinsey Global Institute analysis

BIAS THAT LEADS TO UNFAIR OUTCOMES IS A RISK, INCLUDING WHEN ALGORITHMS ARE TRAINED ON BIASED HISTORICAL DATA

Bias in AI may perpetuate and aggravate existing prejudices and social inequalities, affecting already vulnerable populations and potentially amplifying existing cultural prejudices. This bias may come about through problematic historical data, including sample sizes that are not representative or are inaccurate. For example, AI-based risk scoring for criminal justice purposes may be trained on historic criminal data that include biases (for example, where African Americans are unfairly labeled as high risk). As a result, risk scores from AI would continue to perpetuate this bias. The use of an AI-based sentencing solution containing biased data could have a life-changing impact on individuals. Algorithm and data choices can also introduce bias, for example by producing unequal outcomes for certain groups because of weaker statistical correlations for such groups. Some AI applications already highlight large disparities in accuracy depending on the data used for training algorithms; for example, examination of facial-analysis software shows error rates of 0.8 percent for light-skinned men and 34.7 percent for dark-skinned women.[49]

One key source of bias can be poor data quality. An example of this type of data quality issue arises when using data on past employment records to identify future candidates. Since many employers and industries have historically had disparate distributions of gender and race in their corporate hierarchies, using purely historical data to train AI models could perpetuate discrimination. For example, an AI-powered recruiting tool used by one tech company was abandoned recently after several years of trials because it appeared to show systematic bias against women, resulting from patterns in training data from years of hiring history.[50] To counteract such bias, skilled and diverse data science teams need to take into account potential issues in the training data or otherwise sample intelligently from them.

BREACHING PRIVACY OVER PERSONAL INFORMATION COULD CAUSE HARM

Privacy concerns about sensitive personal data are already rife, and the ability to assuage these worries could help public acceptance of widespread AI use, for for-profit as well as for nonprofit ends. The risk here is that information about individuals, such as financial, tax, or health records, could be made accessible through porous AI systems to those without a legitimate need to access them, causing embarrassment and potentially harm. For example, data on movie recommendations and viewing habits combined with movie database data could potentially identify sexual and political information, which in some countries could put individuals in physical or psychological danger.[51] Another example is when a mental health screening solution has access to private information such as emails and health records. The underlying data are already extremely sensitive, and if mental health status becomes public, the exposure could significantly hurt individuals and set back their recovery, given the continuing stigma surrounding mental health issues.

[49] Joy Buolamwini and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification," *Proceedings of Machine Learning Research*, 2018, volume 81.

[50] Jeffrey Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," Reuters, October 10, 2018, reuters.com/article/amazoncom-jobs-automation/rpt-insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSL2N1WP1RO.

[51] Ryan Singel, "Netflix spilled your Brokeback Mountain secret, lawsuit claims," *Wired*, December 17, 2009, wired.com/2009/12/netflix-privacy-lawsuit/.

SAFE USE AND SECURITY ARE ESSENTIAL FOR SOCIAL GOOD USES OF AI

Ensuring that AI applications are safe and responsible for human use is an essential prerequisite for widespread deployment for social aims. Seeking to further social good with technologies that are dangerous for humans would not only run counter to their core mission but could also spark a backlash, given the potentially large numbers of people involved. For technologies that could affect life and well-being, it will be important to have in place safety mechanisms that comply with existing laws and regulations. For example, if AI misdiagnoses patients in hospitals without a safety mechanism in place, particularly if the systems are directly connected to treatment processes, the outcomes could be catastrophic. Another example is in predictive maintenance: if an AI model fails to recognize that a component of a bus or train needs to be replaced, that could result in a major accident. The framework for accountability and liability for harm done by AI is still evolving.

DECISIONS MADE BY COMPLEX AI MODELS WILL NEED TO BECOME MORE READILY EXPLAINABLE

Explaining in human terms the results from large and complex AI models remains one of the key challenges to achieving user and regulatory acceptance.[52] Opening the AI "black box" to show how decisions are made, which factors are decisive and which are not, will be important for social use of AI. This is especially true when working with stakeholders, including NGOs, who require a basic level of transparency of use and will likely want to provide individuals with clear explanations. Not all AI models are complex, and work is under way to design neural network models that are more readily explainable. Ensuring explainability is especially important in use cases relating to any decision making about individuals, and in particular for cases related to justice and criminal identification, when an accused person needs to be able to appeal a decision in a meaningful way. While a black box solution that predicts recidivism may provide accurate results, it could be difficult to interact with and to understand exactly why the solution identified a person as a threat or a potential reoffender. Use of AI models in cases such as identifying tax fraud could be subject to the similar expectations and/or requirements and therefore need a high level of explainability—or run the risk of being considered unusable.

[52] Michael Chui, James Manyika, and Mehdi Miremadi, "What AI can and can't do (yet) for your business," *McKinsey Quarterly*, January 2018.

MITIGATING RISKS

Effective mitigation strategies typically involve "human in the loop" interventions, in which humans are involved in the decision or analysis loop to validate models and double-check results from AI solutions. They may require cross-functional teams, including domain experts, engineers, product managers, user experience researchers, legal professionals, and others, to test, assess, and flag possible unintended consequences. This must be done on an ongoing basis.

Human analysis of data used to train models may be able to identify issues such as bias and lack of representation. Fairness and security "red teams" could carry out solution tests, and in some cases third parties could be brought in to test solutions using an adversarial approach. Some methods of mitigating this bias demonstrated by university researchers include sampling the data with recognition of the inherent bias and creating synthetic data sets based on known statistics. The appropriate solution depends on the specific use case. For example, if the objective of the AI solution is to detect the incidence of stroke in a patient and identify the subtype of stroke, then a promising solution may be to oversample minority class data to overcome initial imbalances.[53] In the case of predictions of sexual assault prevalence based on geography, a synthetic data set may be the better option because available sets of real data may be too small for effective sampling.

Guardrails to prevent users from blindly trusting AI can be put into place. In medicine, for example, misdiagnoses can be devastating to patients, whether through false positive results that cause distress, wrong or unnecessary treatment or surgeries, or, even worse, false negatives, which could lead to patients missing diagnoses until a disease has reached terminal stage. Ensuring education for patients and mandating that a disclaimer be read every time an AI solution gives a result to a patient could be helpful.

Technology may also be able to find some solutions to these challenges, including explainability. For example, nascent approaches to model transparency include local interpretable model-agnostic (LIME) explanations, which attempt to identify those parts of input data on which a trained model relies most heavily to make predictions. Similarly, a system built by DeepMind that recommends treatments for eye diseases provides doctors with an indication of which aspects of a medical scan prompted the diagnosis.

[53] Yizhao Ni et al., "Towards phenotyping stroke: Leveraging data from a large-scale epidemiological study to detect stroke diagnosis," *PLoS ONE*, 2018, Volume 13, Number 2, journals.plos.org/plosone/article?id=10.1371/journal.pone.0192586.

6. SCALING UP THE USE OF AI FOR SOCIAL GOOD

The potential for AI to be used for social good is compelling, given both the large numbers of people who could be helped and the otherwise intractable problems that may be solved. As with any technology deployment for social good, the scaling up and successful application of AI for societal benefit will depend on the willingness of relevant groups of stakeholders, including collectors and generators of data as well as governments and NGOs, to engage. These are still early days for AI deployment for social use, and considerable progress will be needed before the vast potential becomes a reality.

For now, AI capabilities are being tested and deployed, and they already show promise across a range of domains. The technology itself is advancing rapidly. As we have sought to describe in this paper, the next steps will need to focus on scaling up AI solutions and overcoming the bottlenecks and market failures that are holding it back for now.

Public- and private-sector players all have a role to play. Industries that control or generate data, including satellite and telecommunications companies and major tech firms with social media or other platforms, can accelerate their involvement. That in turn will require agreements on how to handle the cross-collaboration with data in a way that is protective of users and takes into account ethics concerns. NGOs and other social-sector organizations will also need to raise their ability to work with AI solutions. All this leaves ample room for philanthropists and others to help build the capabilities needed to increase AI deployment for social good.

In this final chapter, we suggest some areas in which stakeholders could make a meaningful contribution to further the use of AI for the benefit of society, especially in overcoming the key impediments of data accessibility, talent, and implementation.

IMPROVING DATA ACCESSIBILITY FOR SOCIAL IMPACT CASES WILL REQUIRE ACTIVE PARTICIPATION OF DATA COLLECTORS AND GENERATORS

A wide range of stakeholders owns, controls, collects, or generates the data that could be deployed for AI solutions. Governments are among the most significant collectors of information, which can include tax, health, and education data, but private companies—including satellite operators, telecommunications firms, utilities, and technology companies that run digital platforms, as well as social media sites and search operations—also collect massive volumes of data. These data sets may contain highly confidential personal data that cannot be shared without being anonymized. But, in the case of private operators, data sets may also be monetized by their collectors, and thus not available for pro-bono social good cases.

Overcoming this accessibility challenge will likely require a call to make specific data sets more readily available for well-defined societal initiatives.

Data collectors and generators will need to be encouraged, and possibly mandated, to open access to a subset of their data when it could be in the clear public interest. Such dual use is already starting to happen in some areas, and new business models are being tested that may facilitate data sharing. For example, many satellite data companies participate in the International Charter on Space and Major Disasters, which commits them to open access to satellite data during emergencies such as the September 2018 tsunami in Indonesia and Hurricane Michael, which hit the US East Coast in October 2018.[54] Other data sharing initiatives with private companies are also being worked on, including OPAL, an open

[54] See disasterscharter.org. For other examples of data collaboration, see Stefaan G. Verhulst and Andrew Young, "How the data that Internet companies collect can be used for the public good," *Harvard Business Review*, January 23, 2018.

algorithm collaboration between the World Economic Forum, MIT Media Lab, Orange, and others to derive aggregated insights from a company's data without data leaving the company's server. If proven successful, this could be a powerful tool in unlocking private data for social causes. It started in 2017 with pilots in Colombia and Senegal in partnership with their governments and two telecommunications operators.[55]

Close collaboration between NGOs and data collectors and generators could also help to facilitate this push to make data more accessible. Funding will be required from governments and foundations for initiatives to record and store data that could be used for social ends.

Even if the data are to become accessible, using them presents challenges. Continued investment will be needed to support high-quality data labeling. And multiple stakeholders will need to commit to storing data that can be accessed in a coordinated way and to use the same data standards where possible to ensure seamless interoperability.

Where data sets are freely available, moreover, the data may not have sufficiently large volume for deep learning, which will restrict application of these capabilities—although with advances in transfer learning and pretrained models, some capabilities may not need data volumes as large as would previously have been the case.

Issues of data quality as well as potential bias and fairness will also need to be addressed if the data are to be deployed usefully. Transparency will be a key for the latter issues. A deep understanding of the data, their provenance, and their characteristics will need to be captured so that others using the data set are aware of potential flaws.

This is a long list, and it is likely to require collaboration among companies, governments, and NGOs to set up regular data forums in each industry to work on data availability, accessibility, and connectivity issues, ideally setting global industry standards and collaborating closely on use cases, to ensure that implementation becomes feasible.

OVERCOMING AI TALENT SHORTAGES IS ESSENTIAL FOR SOLVING TECHNICAL CHALLENGES AND IMPLEMENTING AI SOLUTIONS

While solving data challenges will require goodwill and careful coordination, the talent-related challenges we have identified with respect to applying AI for societal good are potentially long-term issues that start with changes in education systems.

The talent challenge is twofold: a shortage of workers with high-level AI expertise—including (but not limited to) PhD-level experience—who are able to develop and train more complex models and a lack of data scientists, translators, and other AI practitioners who can become involved in the implementation phase. The social sector has an especially difficult challenge in hiring and retaining both types of talent, given the high salaries that experienced practitioners can earn at commercial companies.

The long-term solution to these challenges will be to recruit more students to be trained in AI. That could be spurred by significant increases in funding, both grants and scholarships, for tertiary education and PhDs in AI-related fields. With the high salaries that AI expertise commands today, the employment market has signaled a surge in demand for such education.

Sustaining and increasing current training opportunities would be helpful. They include "AI residencies"—one-year training programs at corporate research labs (such as OpenAI, Google, Facebook, and Microsoft)—as well as shorter-term AI boot camps and academies for midcareer professionals. An advanced degree typically is not a prerequisite for these

[55] See opalproject.org/about-opal.

programs, which can train participants in the practice of AI research without having to spend years in a PhD program.

In the absence of experienced professionals in the social sector, companies with AI talent could play a major role in focusing more efforts on AI solutions that have a social impact. For example, they could encourage employees to volunteer and support or coach noncommercial organizations to adopt, deploy, and sustain high-impact AI solutions.

Companies and universities with AI talent could also allocate some of their research capacity to new AI capabilities or solutions that focus on societal benefit but are unable to attract people with the requisite skills.

Overcoming the shortage of talent able to manage AI implementations will likely require government and educational providers to work with companies and social-sector organizations to develop more free or low-cost online training courses. The courses would help those with a basic understanding of computer science acquire the skills needed to pull the data together and be on the frontline of AI implementation (see Box 3, "Ten steps to AI deployment for social good, and some barriers to overcome: A checklist"). Foundations could provide funding for such initiatives.

One idea is to create prizes and other programs to spur innovation for creative solutions. Incentives applied to social good may accelerate the learning process for NGOs, and monetary prizes can encourage new entrants to the field.

Task forces of tech and business translators from government, corporations, the freelance ranks, and social organizations could be established to help teach NGOs about AI with relatable case studies. Beyond coaching, these task forces could potentially help NGOs scope potential projects, support deployment, and plan sustainable road maps. As with all technology deployment, success or failure can depend on the existence of necessary infrastructure as well as the organization's ability to adapt and raise sufficient funding. While these and other implementation challenges are not directly AI-related, they should not be overlooked when attempting to deploy AI for social good.

•••

AI could be the moon landing of our generation, an inspiring scientific leap forward that brings huge benefits to mankind. From the modest library of use cases that we have begun to compile, we can already see tremendous potential to address the world's most important challenges, from hunger to disease to education. Hundreds of millions of people could benefit, from autistic children to earthquake survivors. While the potential to do good with AI is impressive, turning that potential into reality on the scale it deserves will require focus, collaboration, goodwill, funding, and a determination among many stakeholders to work for the benefit of society. Many gaps remain. Some are technological and can be overcome if recent rapid scientific breakthroughs continue. Others relate to talent and the shortage of humans who can develop and train these systems and make them work on the ground. We are only just setting out on this journey. Reaching the destination will be a step-by-step process of confronting barriers and obstacles. We can see the moon—but getting there will require more work and a solid conviction that the goal is worth all the effort, for the sake of everyone.

Box 3. Ten steps to AI deployment for social good, and some barriers to overcome: A checklist

This is a checklist for social-sector organizations thinking about deploying AI solutions, including for the first time. It highlights both the actions that may smooth the path to a successful deployment and the barriers that will need to be overcome. While there can be entry points anywhere along this list, the social-sector organization will need to address steps 1, 2, and 3 first, and then the remaining items on the checklist, to ensure that AI can be deployed.

☑ **1.** Societal problem clearly defined: Specific problem in a social impact domain identified, with measurable objective and requirements for success.

☑ **2.** Translation into technical problem achieved: Formulation of technical problem structure with defined data types and evaluation of technical feasibility completed by technical experts.

☑ **3.** AI technology and/or data confirmed to be a bottleneck in the societal problem and decision to deploy AI (or not) made: Comparison of value of AI to other solutions (for example, policy or changing stakeholder incentives) and evaluation of potential risks and mitigations completed to determine whether AI deployment is the right solution.

- Related potential barriers to overcome: organization receptiveness, access to technology for users, organization deployment efficiency, regulatory limitations.
- Related potential risks to consider and mitigate: data and model bias, data privacy violations, unsafe use of solution, inability to meet explainability level required.

☑ **4.** Organization to deploy solution committed: Social-sector organization with resources to deploy AI solution at scale has committed to solving the defined societal problem.

- Related potential barriers to overcome: organization receptiveness, access to technology for users, organization deployment efficiency, regulatory limitations.

☑ **5.** Required data set available or can be generated: The required data exists in some form, and the data set can feasibly be put together.

- Related potential barriers to overcome: data availability, data integration, data volume.
- Related potential risks to consider and mitigate: data and model bias, data privacy violations.

☑ **6.** Required data set can be accessed: Data released for public use or to the solution builder.

- Related potential barrier to overcome: data accessibility.
- Related potential risks to consider and mitigate: data and model bias, data privacy violations.

☑ **7.** Required data quality ensured and can be used: Data preprocessed into standard format that can be readily fed to a training algorithm.

- Related potential barriers to overcome: data quality, data volume.
- Related potential risks to consider and mitigate: data and model bias, data privacy violations.

☑ **8.** AI model built and trained on available data set. Model proves significant utility on test set.

- Related potential barriers to overcome: availability and accessibility of talent with high-level AI expertise as well as AI practitioners, access to computing capacity.
- Related potential risks to consider and mitigate: unsafe use of solution, inability to meet explainability level required.

☑ **9.** Trained AI model proved value on the ground: Model used in the target environment and demonstrated sufficient value to drive large-scale adoption by actor (for example, NGO).

- Related potential barriers to overcome: availability and accessibility of talent with high-level AI expertise as well as AI practitioners, access to technology for users, organization receptiveness, organization deployment efficiency, regulatory limitations.
- Related potential risks to consider and mitigate: unsafe use of solution, inability to meet explainability level required.

☑ **10.** Organization has built required technical capabilities: Committed organization or organizations have hired or trained required technical capabilities to run and maintain the AI model independently and sustainably.

- Related potential barriers to overcome: AI practitioner availability and accessibility, organization deployment efficiency.

ACKNOWLEDGMENTS

This "Notes from the AI frontier" discussion paper is part of an ongoing series of publications that explores aspects of artificial intelligence and its potential impact on business, the economy, and society. Previous papers have looked at AI use cases across sectors and business functions and AI's potential impact on the global economy.[56]

The research was led by Michael Chui, a McKinsey Global Institute partner in San Francisco; Martin Harrysson, a McKinsey partner in Silicon Valley; James Manyika, chairman and director of the McKinsey Global Institute and McKinsey senior partner based in San Francisco; and Roger Roberts, a McKinsey partner based in Silicon Valley. Ashley van Heteren provided guidance and support. Rita Chung headed the working team, which comprised Tara Balakrishnan, Alexandre Kleis, Pieter Nel, and Sigberto Alarcon Viesca.

We are grateful to colleagues within McKinsey who provided valuable advice and analytical support: Shannon Bouton, Stephanie Carlton, Alberto Chaia, Bertil Chappuis, Ben Cheatham, Kevin Chao, Michael Conway, Matt Craven, Tyler Duvall, Niklas Garemo, Taras Gorishnyy, Lars Hartenstein, Kimberley Henderson, Solveigh Hieronimus, Tarek Mansour, Mekala Krishnan, Marc Krawitz, Pankal Kumar, Dave Levin, Sacha Litman, Anu Madgavkar, Dany Matar, Mona Mourshed, Jan Mischke, Keith Otis, Sangeeth Ram, Chloe Rivera, Matt Rogers, Ben Safran, Hamid Samandari, Dhaval Shah, Jake Silberg, Vivien Singer, Sarah Tucker, Nathan Uhlenbrock, Elke Uhrmann-Klingen, Jonathan Usuka, Helga Vanthournout, Judy Wade, Kate Whittington, and Jonathan Woetzel.

This independent MGI initiative is based on our own research, the experience of our McKinsey colleagues more broadly, and the McKinsey High Tech Practice's research collaboration with Google.org and AI researchers including at Google AI. In particular, we are grateful to Greg Corrado, Kat Chou, Tomas Izo, and Daphne Luong of Google AI, and Micah Berman, Charina Chou, Andrew Dunckelman, Jacquelline Fuller, Brigitte Hoyer Gosselink, Mollie Javerbaum, and Justin Steele at Google.org.

Many others informed our research. We wish to thank: Aaron Horowitz and Lucia Tian at ACLU; Tess Posner at AI4All; James Hodson at AI for Good Foundation; Tom Fairburn and Ben Wylie at The Baobab Network; Paul Duan at Bayes Impact; Amy Guggenheim Shenkan at Common Sense Media; Dean Ramayya Krishnan at Carnegie Mellon University; Jake Porway at Datakind; Paul van der Boor and Rayid Ghani at Data Science for Social Good; Jim Bildner at DRK Foundation; Jeremy Howard at fast.ai; Amy Luers at Future Earth; Dave Levin at Hala Systems; Anju Khetan at Khan Academy; Alan Donald at Mercy Corps; Professor Sandy Pentland at MIT; Will Marshall, Trevor Hammond, and Tara O'Shea at Planet Labs; Jeremy Pierotti at Sansoro Health; Prem Ramaswami at Sidewalk Labs; Professor Marshall Burke at Stanford University; Professor Toby Walsh at University of New South Wales; Professor Milind Tambe at University of Southern California; and Oliver Kharraz at Zocdoc.

This discussion paper was edited and produced by Peter Gumbel, MGI's editorial director, editorial production manager Julie Philpot, senior designers Marisa Carder, Margo Shimasaki, and Patrick White, and Rich Johnson, data visualization editor. Rebeca Robboy, MGI director of external communications, managed dissemination and publicity, while digital editor Lauren Meling provided support for online and social media treatments.

This report contributes to MGI's mission to help business and policy leaders understand the forces transforming the global economy, identify strategic locations, and prepare for the next wave of growth. As with all MGI research, this work is independent, reflects our own views, and has not been commissioned by any business, government, or other institution. We welcome your comments on the research at **MGI@mckinsey.com**.

[56] *Notes from the AI frontier: Insights from hundreds of use cases*, McKinsey Global Institute, April 2018; *Notes from the AI frontier: Modeling the impact of AI on the world economy*, McKinsey Global Institute, September 2018.

RELATED MGI AND MCKINSEY RESEARCH

Notes from the AI frontier: Insights from hundreds of use cases (April 2018)
This paper provides an analysis of more than 400 use cases across 19 industries and nine business functions and highlights the broad use and significant economic potential of advanced AI techniques.

Notes from the AI frontier: Modeling the impact of AI on the world economy (September 2018)
Artificial intelligence has large potential to contribute to global economic activity. But widening gaps among countries, companies, and workers will need to be managed to maximize the benefits.

Artificial intelligence: The next digital frontier? (June 2017)
This paper discusses AI and how companies new to the space can learn a great deal from early adopters who have invested billions into AI and are now beginning to reap a range of benefits.

Jobs lost, jobs gained: Workforce transitions in a time of automation (December 2017)
This report explains that as many as 375 million workers around the world may need to switch occupational categories and learn new skills.

A future that works: Automation, employment, and productivity (January 2017)
This report explores the potential uses of automation in multiple industries and its effects on employment and productivity.

The age of analytics: Competing in a data-driven world (December 2016)
Big data's potential just keeps growing. Taking full advantage means companies must incorporate analytics into their strategic vision and use it to make better, faster decisions.

www.mckinsey.com/mgi
Download and listen to MGI podcasts on iTunes or at www.mckinsey.com/mgi/publications/multimedia/

McKinsey Global Institute
December 2018
Copyright © McKinsey & Company
www.mckinsey.com/mgi

@McKinsey_MGI
McKinseyGlobalInstitute